

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:25:00 ; Search time 18.0843 Seconds  
(without alignments)  
582.095 Million cell updates/sec

Title: US-09-234-208B-1

Perfect score: 418  
Sequence: 1 GTHSLPRPAVPLRMQP.....VGRGPPDAHVAVNLSREY 79

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/genesqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	418	100.0	79	22	AAE09196 Human p68HER-2 ECD
2	418	100.0	83	22	AAE09185 Human p68HER-2 par
3	418	100.0	84	23	AAE20349 HER2 protein conta
4	413	98.8	79	22	AAE09188 Human p68HER-2 ECD
5	411	98.3	79	22	AAE09195 Human p68HER-2 ECD
6	411	98.3	79	22	AAE09199 Human p68HER-2 ECD
7	406	97.1	79	21	AAE09184 HER-2 C-terminal e
8	406	97.1	79	22	AAE09184 Human p68HER-2 ECD
9	406	97.1	79	23	AAE20350 Human HER2 intron
10	404	96.7	79	22	AAE09192 Human p68HER-2 ECD

11	403	96.4	79	22	AAE09190 Human p68HER-2 ECD
12	402	96.2	79	22	AAE09186 Human p68HER-2 ECD
13	400	95.7	79	22	AAE09189 Human p68HER-2 ECD
14	399	95.5	79	22	AAE09187 Human p68HER-2 ECD
15	399	95.5	79	22	AAE09191 Human p68HER-2 ECD
16	398	95.2	79	22	AAE09197 Human p68HER-2 ECD
17	398	95.2	79	22	AAE09198 Human p68HER-2 ECD
18	397	95.0	79	22	AAE09193 Human p68HER-2 ECD
19	396	94.7	79	22	AAE09194 Human p68HER-2 ECD
20	363	86.8	419	22	AAE09213 Human p68HER-2 gen
21	358	85.6	419	22	AAE09212 Human p68HER-2 gen
22	356	85.2	419	22	AAE09205 Human p68HER-2 gen
23	355	84.9	79	21	AAE09239 HER-2 C-terminal p
24	355	84.9	79	22	AAE09180 Human p68HER-2 ECD
25	355	84.9	79	23	AAE20347 Human p68HER-2 ECD
26	354	84.7	419	22	AAE09207 Human p68HER-2 gen
27	354	84.7	419	22	AAE09209 Human p68HER-2 gen
28	352	84.2	419	22	AAE09203 Human p68HER-2 gen
29	351	84.0	419	22	AAE09181 Human p68HER-2 gen
30	351	84.0	419	22	AAE09208 Human p68HER-2 gen
31	351	84.0	419	22	AAE09210 Human p68HER-2 gen
32	351	84.0	419	23	AAE20348 Human truncated HE
33	351	84.0	420	21	AAE09240 Human p68HER-2 gen
34	350	83.7	419	22	AAE09206 Human p68HER-2 gen
35	350	83.7	419	22	AAE09211 Human p68HER-2 gen
36	349	83.5	419	22	AAE09204 Human p68HER-2 gen
37	346	82.8	419	22	AAE09216 Human p68HER-2 gen
38	343	82.1	79	22	AAE09182 Human p68HER-2 ECD
39	343	82.1	419	22	AAE09200 Human p68HER-2 gen
40	343	82.1	419	22	AAE09202 Human p68HER-2 gen
41	339	81.1	419	22	AAE09183 Human p68HER-2 gen
42	337	80.6	419	22	AAE09214 Human p68HER-2 gen
43	337	80.6	419	22	AAE09215 Human p68HER-2 gen
44	86	20.6	459	20	AAE09215 Human p68HER-2 gen
45	86	20.6	459	20	AAE09215 Human p68HER-2 gen

#### ALIGNMENTS

RESULT 1  
AAE09196  
AAE09196 standard; peptide: 79 AA.  
ID  
AC AAE09196;  
XX  
DF 15-NOV-2001 (first entry)  
XX  
DE Human p68HER-2 ECDIIIA variant 11 encoded by HER-2 intron 8.  
XX  
KM HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;  
KW solid tumour; cancer; polymorphism; cytostatic; gene therapy;  
XX p68HER-2; ECDIIIA; variant.  
XX  
XX OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FH  
FT Misc-difference 6 /note="p68HER-2 ECDIIIA (AAE09184) Pro substituted  
FT FT with Leu"  
FT FT  
FT Misc-difference 73 /note="p68HER-2 ECDIIIA (AAE09184) Asp substituted  
FT FT with Asn which is encoded by CAC"

WO200161356-A1.  
23-AUG-2001.  
16-FEB-2001; 2001WO-US05327.  
16-FEB-2000; 2000US-0506079.  
(UOR-) UNIV OREGON HEALTH SCI.

```

XX Clinton G, Henner WD, Evans A;
PI
XX
XX MPI: 2001-529934/58.
DR
XX N-PSDB: AAD15865.
XX
PT New polypeptide, which binds to the extracellular domain of HER-2 for
PT the treatment of hard tumors -
XX
XX Example 11: Page -; 61pp: English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Hereceptin, at an affinity of at least 10-8. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is p68HER-2 ECDIIIA variant
CC Note: The present sequence is not shown in the specification but is
CC derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
CC (AAE09184).
XX
XX Sequence 79 AA:
SQ
Query Match 100.0%; Score 418; DB 22; Length 79;
Best Local Similarity 100.0%; Pred. No. 2e-38;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTHSLPRPAANVPVLRMQPGPAHPVLSFLRPSWDLVSFAFYSLPLAPLSPTSPVPSV 60
DB 1 GTHSLPRPAANVPVLRMQPGPAHPVLSFLRPSWDLVSFAFYSLPLAPLSPTSPVPSV 60
QY 61 GRGPPDAHVAVNLRYEG 79
DB 61 GRGPPDAHVAVNLRYEG 79

RESULT 2
AAE09185
ID AAE09185 standard: Protein; 83 AA.
AC
XX AAE09185;
XX
XX 15-NOV-2001 (first entry)
XX
DE Human p68HER-2 partial protein containing ECDIIIA variant sequence.
XX
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;
XX solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA; variant.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 3..81
FT Domain /label= ECDIIIA
FT /note= "Extracellular domain IIRA of p68HER-2"
FT Misc-difference 1 /note= "Encoded by CC"
FT Misc-difference 8 /note= "Most commonly occurring ECDIIIA (AAE09184) PRO
FT substituted with Leu"
FT Modified-site 75
FT /note= "Asn is N-glycosylated. Most commonly occurring
FT ECDIIIA (AAE09184) Asp substituted with Asn"
XX

```

```

PN WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
XX 16-FEB-2000; 2000US-0506079.
XX
XX (UWOR-) UNIV OREGON HEALTH SCI.
XX
XX Clinton G, Henner WD, Evans A;
PI
XX
XX MPI: 2001-529934/58.
DR
XX N-PSDB: AAD15854.
XX
PT New polypeptide, which binds to the extracellular domain of HER-2 for
PT the treatment of hard tumors -
XX
XX Example 1: Fig 1: 61pp: English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Hereceptin, at an affinity of at least 10-8. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is human p68HER-2 partial protein
CC containing ECDIIIA variant sequence.
XX
XX Sequence 83 AA:
SQ
Query Match 100.0%; Score 418; DB 22; Length 83;
Best Local Similarity 100.0%; Pred. No. 2.1e-38;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTHSLPRPAANVPVLRMQPGPAHPVLSFLRPSWDLVSFAFYSLPLAPLSPTSPVPSV 60
DB 3 GTHSLPRPAANVPVLRMQPGPAHPVLSFLRPSWDLVSFAFYSLPLAPLSPTSPVPSV 62
QY 61 GRGPPDAHVAVNLRYEG 79
DB 63 GRGPPDAHVAVNLRYEG 81

RESULT 3
AAE20349
ID AAE20349 standard: Protein; 84 AA.
AC
XX AAE20349;
XX
XX 18-JUN-2002 (first entry)
XX
XX HER2 protein containing extracellular domain (ECDIIIA).
XX
XX Endothelial growth factor receptor; EGFR; tumour; cytostatic; herstatin;
XX HER-2 receptor tyrosine kinase; squamous cell carcinoma; lung; colon;
XX glial cell tumour; cell growth.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
FH 1..2
FT Misc-difference 1 /note= "Encoded by CCCGA"
FT Modified-site 75..77
FT /note= "Asn is N-glycosylated"
FT Misc-difference 81..82
FT /note= "Encoded by GGCTGAGACGGCCCTTCCCCCACCACCCGCCACC
XX

```

```

FT      TCCTCAGTC"
FT      Misc-difference 83..84
FT      /note= "Encoded by TCCT"
PN      WO200214470-A2.
XX      21-FEB-2002.
XX      PD
XX      PF 14-AUG-2001; 2001WO-US25502.
XX      PR 14-AUG-2000; 2000US-0638834.
XX      PA (UIOR-) UNIV OREGON HEALTH SCI.
XX      PI Clinton GM.
XX      PI
XX      DR MPI: 2002-269185/31.
XX      DR N-PSDB: AAD32539.
XX      DR
XX      PT Treating solid tumor characterized by expression of endothelial growth
XX      PT factor receptor, involves administering recombinant herstatin that
XX      PT binds to extracellular domain of the endothelial growth factor receptor
XX      PT
XX      Example 11; Fig 1; 82pp; English.
XX      PS
XX      CC The present invention relates to a method for treating a solid tumour
XX      CC characterised by endothelial growth factor receptor (EGFR) expression.
XX      CC The method involves administering an agent that binds to an extracellular
XX      CC domain (ECD) of EGFR. The invention also relates to a naturally occurring
XX      CC inhibitor of HER-2 receptor tyrosine kinase called herstatin. The co-
XX      CC expression of herstatin with p185HER2 causes a striking reduction in cell
XX      CC growth that corresponds with suppression of p185 autophosphorylation. The
XX      CC method or a pharmaceutical composition is useful for treating a solid
XX      CC tumour (selected from squamous cell carcinoma, lung carcinoma, colon
XX      CC carcinoma and glioma cell tumour) characterised by EGFR expression. The
XX      CC present sequence is HER2 protein containing extracellular domain
XX      CC (ECDIIIA) sequence.
XX      CC
XX      SQ Sequence 84 AA:

Query Match          100.0%; Score 418; DB 23; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.1e-38;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTHSLPPRAAVPVPLRMQPGPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPISPVSV 60
    |||||||
DB 3 GTHSLPPRAAVPVPLRMQPGPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPISPVSV 62
    |||||||
OY 61 GRGPPDAHVAVNLSRYEG 79
    |||||||
DB 63 GRGPPDAHVAVNLSRYEG 81
    |||||||

RESULT 4
AAE09188
ID AAE09188 standard; peptide: 79 AA.
AC AAE09188;
XX
XX 15-NOV-2001 (first entry)
XX
XX Human p68HER-2 ECDIIIA variant 3 encoded by HER-2 intron 8.
XX
XX DE HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
XX KW solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX KW p68HER-2; ECDIIIA; variant.
XX
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Key MISC-difference 6 /note= "p68HER-2 ECDIIIA (AAE09184) Pro substituted
FT

```

```

FT      with Leu"
XX      PN WO200161356-A1.
XX      PD 23-AUG-2001.
XX      PF 16-FEB-2001; 2001WO-US05327.
XX      PR 16-FEB-2000; 2000US-0506079.
XX      PA (UIOR-) UNIV OREGON HEALTH SCI.
XX      PI Clinton G, Henner WD, Evans A;
XX      PI
XX      DR MPI: 2001-529934/58.
XX      DR N-PSDB: AAD15857.
XX      DR
XX      PT New polypeptide, which binds to the extracellular domain of HER-2 for
XX      PT the treatment of hard tumors -
XX      PT
XX      Example 11; Page -: 61pp; English.
XX      PS
XX      CC The invention relates to novel HER-2 (herstatin-2) antagonist
XX      CC particularly a polypeptide that binds to the extracellular domain (ECD)
XX      CC of HER-2 at a site that is different from the binding site of humanised
XX      CC antibody, Herceptin, at an affinity of at least 1078. The present
XX      CC invention is based upon the initial discovery of an alternative HER-2
XX      CC mRNA transcript with 274 bp insert of intron 8. The translation product
XX      CC of the alternative transcript is a truncated HER-2 protein designated
XX      CC p68HER-2 which lacks the transmembrane and intracellular domains of
XX      CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
XX      CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
XX      CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
XX      CC nucleic acids encoding these are useful to treat, diagnose and identify
XX      CC solid tumours. The present sequence is p68HER-2 ECDIIIA variant
XX      CC encoded by polymorphic form of human HER-2 intron 8.
XX      CC Note: The present sequence is not shown in the specification but is
XX      CC derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
XX      CC (AAE09184).
XX      CC
XX      SQ Sequence 79 AA:

Query Match          98.8%; Score 413; DB 22; Length 79;
Best Local Similarity 98.7%; Pred. No. 7e-38;
Matches 78; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTHSLPPRAAVPVPLRMQPGPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPISPVSV 60
    |||||||
DB 1 GTHSLPPRAAVPVPLRMQPGPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPISPVSV 60
    |||||||
OY 61 GRGPPDAHVAVNLSRYEG 79
    |||||||
DB 61 GRGPPDAHVAVNLSRYEG 79
    |||||||

RESULT 5
AAE09195
ID AAE09195 standard; peptide: 79 AA.
AC AAE09195;
XX
XX 15-NOV-2001 (first entry)
XX
XX Human p68HER-2 ECDIIIA variant 10 encoded by HER-2 intron 8.
XX
XX DE HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
XX KW solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX KW p68HER-2; ECDIIIA; variant.
XX
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Key MISC-difference 73
FT

```

```
FT /note= "p68HER-2 ECDIIIA (AAE09184) Asp substituted
FT with Asn which is encoded by CAC"
PN WO200161356-A1.
XX 23-AUG-2001.
PD 16-FEB-2001; 2001WO-US05327.
XX 16-FEB-2000; 2000US-0506079.
PR 16-FEB-2000; 2000US-0506079.
XX (UYOR-) UNIV OREGON HEALTH SCI.
PA Clinton G, Henner WD, Evans A;
PI Clinton G, Henner WD, Evans A;
XX WPI: 2001-529934/58.
DR N-PSDB: AAD15864.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
PT the treatment of hard tumors -
PS
XX Example 11; Page -: 61pp; English.
XX The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Herceptin, at an affinity of at least 1078. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p155HER-2 but contains ECD I, II of the p155HER-2 and the novel ECDIIIA.
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumors. The present sequence is p68HER-2 ECDIIIA variant
CC encoded by polymorphic form of human HER-2 intron 8.
CC Note: The present sequence is not shown in the specification but is
CC derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
CC (AAE09184).
XX
XX Sequence 79 AA:
SQ
Query Match 98.3%; Score 411; DB 22; Length 79;
Best Local Similarity 98.7%; Pred. No. 1.2e-37;
Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GTHSLPPRAAVPVLRMQPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPVSPISPSV 60
DB 1 GTHSLPPRAAVPVLRMQPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPVSPISPSV 60
OY 61 GRGPPDAHVAVNLSTRYEG 79
DB 61 GRGPPDAHVAVNLSTRYEG 79
RESULT 6
AAE09199
ID AAE09199 standard; peptide; 79 AA.
XX
XX AAE09199:
XX 15-NOV-2001 (first entry)
DT
XX Human p68HER-2 ECDIIIA variant 15 encoded by HER-2 intron 8.
DE
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KW solid tumour; cancer; polymorphism; cytostatic; gene therapy;
KM p68HER-2; ECDIIIA; variant.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
```

```
FT Misc-difference 73
FT /note= "p68HER-2 ECDIIIA (AAE09184) Asp substituted
FT with Asn"
PN WO200161356-A1.
XX 23-AUG-2001.
PD 16-FEB-2001; 2001WO-US05327.
XX 16-FEB-2000; 2000US-0506079.
PR 16-FEB-2000; 2000US-0506079.
XX (UYOR-) UNIV OREGON HEALTH SCI.
PA Clinton G, Henner WD, Evans A;
PI Clinton G, Henner WD, Evans A;
XX WPI: 2001-529934/58.
DR N-PSDB: AAD15868.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
PT the treatment of hard tumors -
PS
XX Example 12; Page -: 61pp; English.
XX The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Herceptin, at an affinity of at least 1078. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p155HER-2 but contains ECD I, II of the p155HER-2 and the novel ECDIIIA.
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumors. The present sequence is p68HER-2 ECDIIIA variant
CC encoded by polymorphic form of human HER-2 intron 8.
CC Note: The present sequence is not shown in the specification but is
CC derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
CC (AAE09184).
XX
XX Sequence 79 AA:
SQ
Query Match 98.3%; Score 411; DB 22; Length 79;
Best Local Similarity 98.7%; Pred. No. 1.2e-37;
Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GTHSLPPRAAVPVLRMQPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPVSPISPSV 60
DB 1 GTHSLPPRAAVPVLRMQPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPVSPISPSV 60
OY 61 GRGPPDAHVAVNLSTRYEG 79
DB 61 GRGPPDAHVAVNLSTRYEG 79
RESULT 7
AAE97241
ID AAE97241 standard; Protein; 79 AA.
XX
XX AAE97241:
XX 04-DEC-2000 (first entry)
DT
XX HER-2 C-terminal extracellular domain IIIA.
DE
XX HER-2; erbB-2; oncogene; receptor-like tyrosine kinase; insertion;
KW extracellular domain IIIA; antagonist; intron 8; C-terminal extension;
KM truncated HER-2; p68; dimerization inhibitor; cytostatic.
XX
XX Homo sapiens.
OS
XX
```



PN WO200044403-A1.  
XX  
PD 03-AUG-2000.  
XX  
XX 20-JAN-2000; 2000WO-US01484.  
PF  
XX 20-JAN-1999; 99US-0234208.  
PR  
XX (UYOR-) UNIV OREGON HEALTH SCI.  
PA  
XX Doherty JK, Clinton GM, Adelman JP;  
XX WPI; 2000-499287/44.  
DR N-PSDB; AAA53783.  
XX  
PT using polypeptides and antibodies that bind to the extracellular domain  
of the receptor-like tyrosine kinase HER-2 to treat solid tumors of the  
breast, lung, ovaries and colon  
XX  
XX Example 11; Page 42-43; 46pp; English.  
XX  
XX HER-2/neu (erbB-2) oncogene encodes a receptor-like tyrosine kinase. The  
extracellular domain of p185-HER-2 is proteolytically shed from breast  
carcinoma cells in culture and is found in serum of some cancer patients  
and may be a serum marker of metastatic breast cancer. An alternative  
CC HER-2 mRNA of 4.8 kb with a 274 bp insert (intron 8) has been  
identified. The retained intron is in-frame and encodes a 79 amino acid  
extension designated ECDIIIA (the present sequence), which is inserted at  
residue 340 of p185-HER-2. The alternative mRNA predicts a truncated  
CC HER-2 protein (approximately 68 kDa) that lacks the transmembrane and  
intracellular domains (see AAY97240). p68HER-2 specifically binds to  
CC p185-HER-2 without activating HER-2. It could therefore block  
CC dimerization of p185-HER-2. The p68HER-2 polypeptide binds to a site on  
CC the ECD of HER-2 that is different from the site of binding for  
CC Hereceptin (RTM) (a marketed humanized monoclonal antibody that is used  
for the treatment of cancer and binds to the ECD of HER-2). The methods,  
CC compositions, polypeptides and antibodies are used to treat solid  
CC tumors such as breast cancer, small cell lung carcinoma, ovarian cancer  
and/or colon cancer, especially where over-expression of HER-2 is  
indicated.  
XX  
SQ Sequence 79 AA:  
Query Match 97.1%; Score 406; DB 21; Length 79;  
Best Local Similarity 97.5%; Pred. No. 4.1e-37;  
Matches 77; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTHSLPRRAAVPVPLRMQPGRAHPVLSFLRPSWDLVSFAFYSLLPLAPLSPTSPVSPSV 60  
DB 1 GTHSLPRRAAVPVPLRMQPGRAHPVLSFLRPSWDLVSFAFYSLLPLAPLSPTSPVSPSV 60  
QY 61 GRGPPDDAHVAANLSRYEG 79  
DB 61 GRGPPDDAHVAANLSRYEG 79  
RESULT 8  
AAE09184  
ID AAE09184 standard; peptide; 79 AA.  
XX  
AC AAE09184;  
XX  
DT 15-NOV-2001 (first entry)  
XX  
DE Human p68HER-2 ECDIIIA domain encoded by HER-2 intron 8.  
XX  
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;  
KW solid tumour; cancer; polymorphism; cytostatic; gene therapy;  
KM p68HER-2; ECDIIIA; intron 8.  
XX  
OS Homo sapiens.  
XX  
PN WO200161356-A1.

XX 23-AUG-2001.  
XX  
PD 16-FEB-2001; 2001WO-US05327.  
XX  
PF 16-FEB-2000; 2000US-0506079.  
XX  
XX 16-FEB-2000; 2000US-0506079.  
PR  
XX (UYOR-) UNIV OREGON HEALTH SCI.  
PA  
XX Clinton G, Henner WD, Evans A;  
XX WPI; 2001-529934/58.  
DR N-PSDB; AAD15853, AAD15869.  
XX  
PT New polypeptide, which binds to the extracellular domain of HER-2 for  
the treatment of hard tumors -  
XX  
XX Example 11; Fig 8; 61pp; English.  
XX  
XX The invention relates to novel HER-2 (herstatin-2) antagonist  
CC particularly a polypeptide that binds to the extracellular domain (ECD)  
of HER-2 at a site that is different from the binding site of humanised  
CC antibody, Hereceptin, at an affinity of at least 10<sup>-8</sup>. The present  
CC invention is based upon the initial discovery of an alternative HER-2  
CC mRNA transcript with 274 bp insert of intron 8. The translation product  
of the alternative transcript is a truncated HER-2 protein designated  
CC p68HER-2 which lacks the transmembrane and intracellular domains of  
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.  
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise  
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the  
CC nucleic acids encoding these are useful to treat, diagnose and identify  
solid tumors. The present sequence is ECDIIIA domain of human p68HER-2  
protein encoded by intron 8 of HER-2 gene.  
XX  
SQ Sequence 79 AA:  
Query Match 97.1%; Score 406; DB 22; Length 79;  
Best Local Similarity 97.5%; Pred. No. 4.1e-37;  
Matches 77; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTHSLPRRAAVPVPLRMQPGRAHPVLSFLRPSWDLVSFAFYSLLPLAPLSPTSPVSPSV 60  
DB 1 GTHSLPRRAAVPVPLRMQPGRAHPVLSFLRPSWDLVSFAFYSLLPLAPLSPTSPVSPSV 60  
QY 61 GRGPPDDAHVAANLSRYEG 79  
DB 61 GRGPPDDAHVAANLSRYEG 79  
RESULT 9  
AAE20350  
ID AAE20350 standard; Protein; 79 AA.  
XX  
AC AAE20350;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Human HER2 intron 8 encoded protein.  
XX  
XX Human; tumour; endothelial growth factor receptor; EGFR; cytostatic;  
KW herstatin; HER-2 receptor tyrosine kinase; squamous cell carcinoma; lung;  
KM colon; glial cell tumour; cell growth.  
XX  
OS Homo sapiens.  
XX  
PN WO200214470-A2.  
XX  
PD 21-FEB-2002.  
XX  
PF 14-AUG-2001; 2001WO-US25502.  
XX  
XX 14-AUG-2000; 2000US-0638834.  
XX  
PN

```
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Clinton GW.
PI
XX WPI: 2002-269185/31.
XX DR N-PSDB: AAD32540.
XX
XX Treating solid tumor characterized by expression of endothelial growth
PT factor receptor, involves administering recombinant herstatin that
PT binds to extracellular domain of the endothelial growth factor receptor
PT
XX
XX Example 11: Fig 8: 82pp: English.
XX
XX The present invention relates to a method for treating a solid tumour
CC characterised by endothelial growth factor receptor (EGFR) expression.
CC The method involves administering an agent that binds to an extracellular
CC domain (ECD) of EGFR. The invention also relates to a naturally occurring
CC inhibitor of HER-2 receptor tyrosine kinase called herstatin. The co-
CC expression of herstatin with p185HER2 causes a striking reduction in cell
CC growth that corresponds with suppression of p185 autophosphorylation. The
CC method or a pharmaceutical composition is useful for treating a solid
CC tumour (selected from squamous cell carcinoma, lung carcinoma, colon
CC carcinoma and glial cell tumour) characterised by EGFR expression. The
CC present sequence is a protein encoded by human HER2 intron 8.
XX
XX Sequence 79 AA:
SQ
XX
XX Query Match 97.1%; Score 406; DB 23; Length 79;
XX Best Local Similarity 97.5%; Pred. No. 4.1e-37;
XX Matches 77; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GTHSLPRRAAVVPLRMQPGRAHPVLSFLRPSMDLSAFYSPLAPLSPTSPVSPSV 60
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 1 GTHSLPRRAAVVPLRMQPGRAHPVLSFLRPSMDLSAFYSPLAPLSPTSPVSPSV 60
XX
XX QY 61 GRGPPDAHAHVANLSRYEG 79
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 61 GRGPPDAHAHVANLSRYEG 79
XX
XX RESULT 10
XX AAE09192
XX ID AAE09192 standard; peptide: 79 AA.
XX
XX AC AAE09192;
XX
XX DT 15-NOV-2001 (first entry)
XX
XX DE Human p68HER-2 ECDIIIA variant 7 encoded by HER-2 intron 8.
XX
XX KW HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;
XX solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA; variant.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 36 /note="p68HER-2 ECDIIIA (AAE09184) Leu substituted
XX FT with Ile"
XX
XX PN WO200161356-A1.
XX
XX PD 23-AUG-2001.
XX
XX PF 16-FEB-2001; 2001WO-US05327.
XX
XX PR 16-FEB-2000; 2000US-0506079.
XX
XX PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX PI Clinton G, Henner WD, Evans A;
```

```
XX
XX WPI: 2001-529934/58.
XX DR N-PSDB: AAD15861.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
PT the treatment of hard tumors -
PT
XX
XX Example 11: Page -: 61pp: English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Hereptin, at an affinity of at least 10-8. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is p68HER-2 ECDIIIA variant
CC encoded by polymorphic form of human HER-2 intron 8.
CC Note: The present sequence is not shown in the specification but is
CC derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
CC (AAE09184).
XX
XX Sequence 79 AA:
SQ
XX
XX Query Match 96.7%; Score 404; DB 22; Length 79;
XX Best Local Similarity 96.2%; Pred. No. 6.8e-37;
XX Matches 76; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GTHSLPRRAAVVPLRMQPGRAHPVLSFLRPSMDLSAFYSPLAPLSPTSPVSPSV 60
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 1 GTHSLPRRAAVVPLRMQPGRAHPVLSFLRPSMDLSAFYSPLAPLSPTSPVSPSV 60
XX
XX QY 61 GRGPPDAHAHVANLSRYEG 79
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 61 GRGPPDAHAHVANLSRYEG 79
XX
XX RESULT 11
XX AAE09190
XX ID AAE09190 standard; peptide: 79 AA.
XX
XX AC AAE09190;
XX
XX DT 15-NOV-2001 (first entry)
XX
XX DE Human p68HER-2 ECDIIIA variant 5 encoded by HER-2 intron 8.
XX
XX KW HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;
XX solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA; variant.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 18 /note="p68HER-2 ECDIIIA (AAE09184) Met substituted
XX FT with Leu which is encoded by A7A"
XX
XX PN WO200161356-A1.
XX
XX PD 23-AUG-2001.
XX
XX PF 16-FEB-2001; 2001WO-US05327.
XX
XX PR 16-FEB-2000; 2000US-0506079.
XX
XX PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX
```

PI	ClIntlong; Henner MD, Evans A;
XX	
DR	WPI: 2001-529934/58.
XX	N-PsDB: AAD15859.
XX	
PT	New polypeptide, which binds to the extracellular domain of HER-2 for
PT	the treatment of hard tumors -
PS	Example 11: Page -, 61pp; English.
CC	The invention relates to novel HER-2 (herstatin-2) antagonist
CC	particularly a polypeptide that binds to the extracellular domain (ECD)
CC	of HER-2 at a site that is different from the binding site of humanised
CC	antibody, Herceptin, at an affinity of at least 10 <sup>-8</sup> . The present
CC	invention is based upon the initial discovery of an alternative HER-2
CC	mRNA transcript with 274 bp insert of intron 8. The translation product
CC	of the alternative transcript is a truncated HER-2 protein designated
CC	p68HER-2 but lacks the transmembrane and intracellular domains of
CC	p185HER-2 but contains ECD 1, II of the p185HER-2 and the novel ECDIIIA.
CC	The ECDIIIA-containing polypeptides bind tightly to an, and thus antagonise
CC	the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC	nucleic acids encoding these are useful to treat, diagnose and identify
CC	solid tumours. The present sequence is p68HER-2 ECDIIIA variant
CC	encoded by polymorphic form of human HER-2 intron 8.
CC	Note: The present sequence is not shown in the specification but is
CC	derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
CC	(AAE09184).
SQ	Sequence 79 AA:
	Query Match 96.4%; Score 403; DB 22; Length 79;
	Best Local Similarity 96.2%; Pred. No. 8.7e-37;
	Matches 76; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY	1 GTHTLPPRAAPVPLRMOPGPAHVLSLRPSMWLVSAFYSLPLAIPSTSPVSPVSV 60
Db	1 GTHTLPPRAAPVPLRLQPGPAHVLSLRPSMDLVSAFYSLPLAIPSTSPVSPVSV 60
OY	61 GRGPDDAHVAVNLSRYEG 79
Db	61 GRGPDDAHVAVDLSRYEG 79
	RESULT 12
	AAE09186
ID	AAE09186 standard; peptide; 79 AA.
AC	
XX	AAE09186;
DT	15-NOV-2001 (first entry)
XX	
DE	Human p68HER-2 ECDIIIA variant 1 encoded by HER-2 intron 8.
XX	
KM	HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KM	solid tumour; cancer; polymorphism; cytostatic; gene therapy;
KW	p68HER-2; ECDIIIA; variant.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 2
FT	/note="p68HER-2 ECDIIIA (AAE09184) Thr substituted
FT	with Ser"
XX	
PB	WO200161356-A1.
PN	
PD	23-AUG-2001.
XX	
PF	16-FEB-2001; 2001WO-US05327.
XX	
PR	16-FEB-2000; 2000US-0506079.
XX	
PA	(UYOR-) UNIV OREGON HEALTH SCI.

XX	Clinton G, Henner WD, Evans A;
PI	
XX	
DR	WPI; 2001-529934/58.
DR	N-PSDB; AAD15855.
XX	
PT	New polypeptide, which binds to the extracellular domain of HER-2 for
PT	the treatment of hard tumors -
XX	
PS	Example 11: Page -: 61pp; English.
XX	
CC	The invention relates to novel HER-2 (herstatin-2) antagonist
CC	particularly a polypeptide that binds to the extracellular domain (ECD)
CC	of HER-2 at a site that is different from the binding site of humanised
CC	antibody, Herceptin, at an affinity of at least 10 <sup>8</sup> . The present
CC	invention is based upon the initial discovery of an alternative HER-2
CC	mRNA transcript with 274 bp insert of intron 8. The translation product
CC	of the alternative transcript is a truncated HER-2 protein designated
CC	p68HER-2 which lacks the transmembrane and intracellular domains of
CC	p185HER-2 but contains ECD 1, II of the p185HER-2 and the novel ECDIIIA.
CC	The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC	the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC	nucleic acids encoding these are useful to treat, diagnose and identify
CC	solid tumours. The present sequence is p68HER-2 ECDIIIA variant
CC	encoded by polymorphic form of human HER-2 intron 8.
CC	Note: The present sequence is not shown in the specification but is
CC	derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
CC	(AAE09184).
XX	
SO	Sequence 79 AA:
Query Match	96.2%; Score 402; DB 22; Length 79;
Best Local Similarity	96.2%; Pred. No. 1,1e-36;
Matches	76; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY	1 GTGSLPRPAVPVPLRMQPGPAHPVLSFLRPSWDLVSAFYSLPLAPLSPTSPVSPSV 60
Db	1 GSHSLPRPAAPVPLRMQPGPAHPVLSFLRPSWDLVSAFYSLPLAPLSPTSPVSPSV 60
OY	61 GRGPPDAHVAANLSRYEG 79
Db	61 GRGPPDAHVAANLSRYEG 79
RESULT 13	
AAE09189	
ID	AAE09189 standard; peptide: 79 AA.
XX	
AC	AAE09189;
XX	
DT	15-NOV-2001 (first entry)
XX	
DE	Human p68HER-2 ECDIIIA variant 4 encoded by HER-2 intron 8.
XX	
KM	HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KM	solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX	p68HER-2; ECDIIIA; variant.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
FT	Misc-difference 16
FT	/note= "p68HER-2 ECDIIIA (AAE09184) Leu substituted
XX	with Gln"
XX	
PN	WC0200161356-A1.
PD	23-AUG-2001.
XX	
XX	16-FEB-2001; 2001WO-US05327.
XX	
XX	16-FEB-2000; 2000US-0506079.
XX	

```
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Clinton G, Henner WD, Evans A;
XX
XX WPI: 2001-529934/58.
XX
XX N-PSDB: AAD15858.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
XX the treatment of hard tumors -
XX
XX Example 11; Page -: 61pp; English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
XX particularly a polypeptide that binds to the extracellular domain (ECD)
XX of HER-2 at a site that is different from the binding site of humanised
XX antibody, Hereceptin, at an affinity of at least 10-8. The present
XX invention is based upon the initial discovery of an alternative HER-2
XX mRNA transcript with 274 bp insert of intron 8. The translation product
XX of the alternative transcript is a truncated HER-2 protein designated
XX p68HER-2 which lacks the transmembrane and intracellular domains of
XX p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
XX The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
XX the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
XX nucleic acids encoding these are useful to treat, diagnose and identify
XX solid tumours. The present sequence is p68HER-2 ECDIIIA variant
XX encoded by polymorphic form of human HER-2 intron 8.
XX Note: The present sequence is not shown in the specification but is
XX derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
XX (AAE09184).
XX
XX SQ Sequence 79 AA:
XX
XX Query Match 95.7%; Score 400; DB 22; Length 79;
XX Best Local Similarity 96.2%; Pred. No. 1.9e-36;
XX Matches 76; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GTHSLPRRAAVVPLRMOPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSVIPSPSV 60
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1 GTHSLPRRAAVVPLRMOPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSVIPSPSV 60
XX
XX QY 61 GRGPPDAHVAVNLRYEG 79
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 61 GRGPPDAHVAVNLRYEG 79
XX
XX DB 61 GRGPPDAHVAVNLRYEG 79
XX
XX RESULT 14
XX AAE09187
XX ID AAE09187 standard; peptide; 79 AA.
XX
XX AC AAE09187;
XX
XX DT 15-NOV-2001 (first entry)
XX
XX DE Human p68HER-2 ECDIIIA variant 2 encoded by HER-2 intron 8.
XX
XX KW HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;
XX solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA; variant.
XX
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 5
XX FT /note="p68HER-2 ECDIIIA (AAE09184) Leu substituted
XX FT with Pro"
XX
XX WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
XX 16-FEB-2000; 2000US-0506079.
XX
```

```
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Clinton G, Henner WD, Evans A;
XX
XX WPI: 2001-529934/58.
XX
XX N-PSDB: AAD15856.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
XX the treatment of hard tumors -
XX
XX Example 11; Page -: 61pp; English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
XX particularly a polypeptide that binds to the extracellular domain (ECD)
XX of HER-2 at a site that is different from the binding site of humanised
XX antibody, Hereceptin, at an affinity of at least 10-8. The present
XX invention is based upon the initial discovery of an alternative HER-2
XX mRNA transcript with 274 bp insert of intron 8. The translation product
XX of the alternative transcript is a truncated HER-2 protein designated
XX p68HER-2 which lacks the transmembrane and intracellular domains of
XX p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
XX The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
XX the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
XX nucleic acids encoding these are useful to treat, diagnose and identify
XX solid tumours. The present sequence is p68HER-2 ECDIIIA variant
XX encoded by polymorphic form of human HER-2 intron 8.
XX Note: The present sequence is not shown in the specification but is
XX derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
XX (AAE09184).
XX
XX SQ Sequence 79 AA:
XX
XX Query Match 95.5%; Score 399; DB 22; Length 79;
XX Best Local Similarity 96.2%; Pred. No. 2.4e-36;
XX Matches 76; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GTHSLPRRAAVVPLRMOPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSVIPSPSV 60
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1 GTHSLPRRAAVVPLRMOPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSVIPSPSV 60
XX
XX QY 61 GRGPPDAHVAVNLRYEG 79
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 61 GRGPPDAHVAVNLRYEG 79
XX
XX DB 61 GRGPPDAHVAVNLRYEG 79
XX
XX RESULT 15
XX AAE09191
XX ID AAE09191 standard; peptide; 79 AA.
XX
XX AC AAE09191;
XX
XX DT 15-NOV-2001 (first entry)
XX
XX DE Human p68HER-2 ECDIIIA variant 6 encoded by HER-2 intron 8.
XX
XX KW HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;
XX solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA; variant.
XX
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 21
XX FT /note="p68HER-2 ECDIIIA (AAE09184) Gly substituted
XX FT with Asp, Ala or Val"
XX
XX WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
XX
```

PR 16-FEB-2000; 2000US-0506079.

XX  
PA (UYOR-) UNIV OREGON HEALTH SCI.  
XX

PI Clinton G, Henner WD, Evans A.

XX WPI; 2001-529934/58.

DR N-PSDB; AAD15860.

XX  
PT New polypeptide, which binds to the extracellular domain of HER-2 for  
XX the treatment of hard tumors -

PS Example 11; Page -: 61pp; English.

XX  
XX The invention relates to novel HER-2 (herstatin-2) antagonist  
CC particularly a polypeptide that binds to the extracellular domain (ECD)  
CC of HER-2 at a site that is different from the binding site of humanised  
CC antibody, Herceptin, at an affinity of at least 10<sup>-8</sup>. The present  
CC invention is based upon the initial discovery of an alternative HER-2  
CC mRNA transcript with 274 bp insert of intron 8. The translation product  
CC of the alternative transcript is a truncated HER-2 protein designated  
CC p68HER-2 which lacks the transmembrane and intracellular domains of  
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.  
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise  
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the  
CC nucleic acids encoding these are useful to treat, diagnose and identify  
CC solid tumours. The present sequence is p68HER-2 ECDIIIA variant  
CC encoded by polymorphic form of human HER-2 intron 8.  
CC Note: The present sequence is not shown in the specification but is  
CC derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8  
CC (AAE09184).

XX  
SQ Sequence 79 AA;

Query Match 95.5%; Score 399; DB 22; Length 79;  
Best Local Similarity 96.2%; Pred. No. 2,4e-36;

Matches 76; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTHSLPPRAAVVPLRMOPGAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSVPTSPVSV 60  
|||||  
Db 1 GTHSLPPRAAVVPLRMOPGAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSVPTSPVSV 60

OY 61 GRGPDPDAHVAVNLRSRYEG 79  
|||||  
Db 61 GRGPDPDAHVAVNLRSRYEG 79

Search completed: March 4, 2003, 12:34:31  
Job time : 19.0843 secs





```

A:Reference number: AB44420; MUID:20083487; PMID:10617197
A:Accession: NB4715
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <STD>
A:Cross-references: GB:AE002093; NID:g3746072; PIDN:AAC63847.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g31050
A:Map position: 2

Query Match
Best Local Similarity      19.4%; Score 81; DB 2; Length 200;
                          34.3%; Pred. No. 0.48;
Matches    23; Conservative   5; Mismatches    23; Indels    16; Gaps    2;

OY      9 PAAYVPLRMOPGPAHPVLFLRPSWDLVSAFYSLLPLAPLSPTSVIPSVSGRPDPDA 68
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       132 PVAAEV-----PGVPRRPSSFSPPS-----QSLPASPVNAHAPVQYOMGPSPAP 175
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      69 HVAVNLS 75
          | | | | |
Db       176 HSAASNS 182

RESULT 3
G86292
hypothetical protein F7H2.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: G86292
R:Theologus, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Hughes, B.; Huizart, L.
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucero, J.S.; Matli, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719; PMID:11130712
A:Accession: G86292
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-1006 <STOD>
A:Cross-references: GB:AE005172; NID:g8927662; PIDN:MAF82153.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match
Best Local Similarity      19.4%; Score 81; DB 2; Length 1006;
                          31.4%; Pred. NO. 3;
Matches    22; Conservative   5; Mismatches    21; Indels    22; Gaps    2;

OY      6 LPRPAVPEPLRMOPGPAHPVLFLRPSWDLVSAFYSLLPLAPLSPTSVIPSVSGRGPD 65
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       336 LPPATRLPPLPLPLPPP-----SLPYPCSP---PPPIIYNAGAP 373
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      66 PDAAHVANLS 75
          | | | | |
Db       374 PCCVTCGVVS 383

RESULT 4
JDM01
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Arabidopsis thaliana
N:Alternate names: DNA-directed RNA polymerase II 205K chain; protein F4B14.70
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 31-Mar-1993 #sequence_revision 11-Jun-1999 #text_change 20-Aug-1999
C:Accession: T04650; S12071; S27346; S11960
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hohenseil, J.; Mewes, H.W.; Mayer, K.F.;
submitted to the Protein Sequence Database, October 1998
A:Reference number: T04650
A:Accession: T04650
A:Molecule type: DNA

```

A:Residues: 1-1834 <REV>  
A:Cross-references: EMBL:AL031986  
A:Experimental source: cultivar Columbia; BAC clone FAB14  
R:Navrath, C.; Schell, J.; Koncz, C.  
Mol. Gen. Genet. 223, 65-75, 1990  
A:Title: Homologous domains of the largest subunit of eucaryotic RNA polymerase II are  
A:Reference number: S12071; MUID:91080867; PMID:2259344  
A:Accession: S12071  
A:Molecule type: DNA  
A:Residues: 1-421,'S','423-732','D','734-1055','R','1057-1714','SPSPSP','1715-1834 <NMM1>  
A:Cross-references: EMBL:X52954; NID:916504; PIDN:CA037130.1; PID:916505  
A:Experimental source: cv. Columbia  
A:Note: the authors translated the codon AGC for residue 1755 as Arg  
A:Accession: S27346  
A:Molecule type: mRNA  
A:Residues: 510-732,'D','734-1055','R','1057-1714','SPSPSP','1715-1834 <NMM2>  
A:Experimental source: cv. Columbia  
R:Dieckrich, M.A.; Prenger, J.P.; Guilfoyle, T.J.  
Plant Mol. Biol. 15, 207-223, 1990  
A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A  
A:Reference number: S11960; MUID:91355869; PMID:2103447  
A:Accession: S11960  
A:Molecule type: DNA  
A:Residues: 1-116,125-192,'NSKEE','198-297','R','299-302','R','304-400','KE','403','VDYGP  
A:Cross-references: EMBL:X52494; NID:916493; PIDN:CA036735.1; PID:916494  
A:Note: the authors translated the codon CCT for residue 1083 as Ala  
C:Genetics:  
A:Gene: RP11215; RPB1  
A:Map position: 4  
A:Introns: 28/3; 123/3; 218/3; 271/3; 325/3; 411/1; 440/3; 552/3; 648/2; 734/3; 1755/  
A:Note: FAB14.70  
C:Superfamily: human DNA-directed RNA polymerase II largest chain  
C:Keywords: DNA binding; nucleotidyltransferase; phosphoprotein; tandem repeat; trans  
F:65-109/Region: zinc finger CCCC motif  
F:1531-1812/Region: 7-residue repeats  
Query Match 19.1%; Score 80; DB 1; Length 1834;  
Best Local Similarity 36.4%; Pred. No. 7.6;  
Matches 28; Conservative 3; Mismatches 36; Indels 10; Gaps 3;  
OY 7 PRPAVPPPLKRPQ-----GPAHPVLSELRPSMDLVSATSLPLAPLSTSPVPSVSGR 62  
DB 1730 PSIAVSPSNARISPA SPSPSPNYSPTSPSPSYSS---PSSPTYSPSPSYSSGA 1785  
OY 63 GPD--PDAAVAVNLRSY 77  
DB 1786 SPDYSPSAGYSPTLPGLY 1802  
RESULT 5  
G85422  
hypothetical protein AT9435800 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence television 16-Feb-2001 #text change 02-Mar-2001  
C:Accession: G85422  
R:anonymus, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: AB5001; MUID:20083488; PMID:10617158  
A:Accession: G85422  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1840 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7270532; PIDN:CAB81489.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT9435800  
A:Map position: 4  
C:Superfamily: human DNA-directed RNA polymerase II largest chain  
Query Match 19.1%; Score 80; DB 2; Length 1840;  
Best Local Similarity 36.4%; Pred. No. 7.7;  
Matches 28; Conservative 3; Mismatches 36; Indels 10; Gaps 3;



QY 7 PRAAVPVPLRMQ-----GPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPVSVGR 62  
 DB 1736 PSIAVPSNMRSLRSPASPSPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSGA 1791  
 QY 63 GPD--PDAAVAVNLSRY 77  
 DB 1792 SPDSPSAGYSPTLPGL 1808

## RESULT 6

S72619

hypothetical protein 1a - anthracnose fungus (Colletotrichum gloeosporioides) retrotrans

C:Species: Glomerella cingulata, Colletotrichum gloeosporioides

C:Date: 24-Oct-1998 #sequence\_revision 24-Oct-1998 #text\_change 17-Mar-1999

C:Accession: S72619

R:He, C.; Nourse, J.P.; Kelemu, S.; Irwin, J.A.G.; Manners, J.M.

Mol. Gen. Genet. 257, 320-331, 1996

A:Title: Cg11: a non-LTR retrotransposon with restricted distribution in the fungal phyt

A:Reference number: S72619; M0ID:96439839; PMID:8842152

A:Accession: S72619

A:Molecule type: DNA

A:Residues: 1-249 &lt;HEA&gt;

A:Cross-references: EMBL:L76172

A:Experimental source: biotype B, isolate U062

A:Note: In the authors' translation residues 1-10 are not shown

C:Genetics:

A:Mobile element: retrotransposon Cg11

Query Match 18.7%; Score 78; DB 2; Length 249;  
 Best Local Similarity 34.7%; Pred. No. 1.2;  
 Matches 25; Conservative 7; Mismatches 26; Indels 14; Gaps 3;  
 QY 5 LPR-----PAVVPVPLRMQGPAPVLSFLRPSW---DLVSIFYSLPLAPLSPTSPV 54  
 DB 24 LVRWVCHPNTNQAQTPRQGRSPAMASQPAAGPSQGLILSSMHNLPKTPPPPTSLP 83  
 QY 55 ISPVSVGRGPPD 66  
 DB 84 PRP-----RGLTP 91

## RESULT 7

T51031

related to finger protein XFG 68 [imported] - Neurospora crassa

N:Alternate names: protein B15120.10

C:Species: Neurospora crassa

C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000

C:Accession: T51031

R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, July 2000

A:Reference number: 225286

A:Accession: T51031

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-217 &lt;SCH&gt;

A:Cross-references: EMBL:AL389900; GSPDB:GN00116; NCSP:B15120.10

A:Experimental source: BAC clone B15120; strain OR74A

C:Genetics:

A:Gene: NCSP:B15120.10

A:Map position: 6

Query Match 18.5%; Score 77.5; DB 2; Length 217;  
 Best Local Similarity 37.5%; Pred. No. 1.2;  
 Matches 21; Conservative 5; Mismatches 21; Indels 9; Gaps 2;  
 QY 7 PRAAVPVPLRMQGPAPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPVSVGR 62  
 DB 109 PRPSTASTPRRCCTPPRP-----PS-----TSALVLLPVSPILPSSSSPILPCSPRR 155

## RESULT 8

T19319

hypothetical protein C15H11.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T19319  
 R:Bardill, S.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19107  
 A:Accession: T19319  
 A>Status: preliminary  
 A:Molecule type: translated from GB/EMBL/DBJ  
 A:Residues: 1-503 <WIL>  
 A:Cross-references: EMBL:281035; PIDN:CAB02737.1; GSPDB:GN00023; CESP:C15H11.5  
 A:Experimental source: clone C15H11  
 C:Genetics:  
 A:Gene: CESP:C15H11.5  
 A:Map position: 5  
 A:Introns: 53/3; 156/3; 254/3; 292/2; 331/3; 379/3; 437/1  
 Query Match 18.4%; Score 77; DB 2; Length 503;  
 Best Local Similarity 32.7%; Pred. No. 3.4;  
 Matches 17; Conservative 8; Mismatches 17; Indels 10; Gaps 1;  
 QY 11 AVPVPLRMQGPAPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPVSVGR 62  
 DB 298 SIPLERMQEPG-----SGWYLSAWYSLPNVPLVMVTCVIRPAEVAR 339

## RESULT 9

T19673

hypothetical protein C33B4.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T19673

R:Coles, L.

submitted to the EMBL Data Library, February 1995

A:Reference number: Z19160

A:Accession: T19673

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1110 &lt;WIL&gt;

A:Cross-references: EMBL:248367; PIDN:CA88324.1; GSPDB:GN00020; CESP:C33B4.3

A:Experimental source: clone C33B4

C:Genetics:

A:Gene: CESP:C33B4.3

A:Map position: 2

A:Introns: 20/3; 110/3; 144/3; 341/1; 455/1; 546/1; 653/3; 934/2; 983/3; 1035/1

Query Match 18.3%; Score 76.5; DB 2; Length 1110;  
 Best Local Similarity 28.6%; Pred. No. 9.6;  
 Matches 24; Conservative 12; Mismatches 33; Indels 15; Gaps 4;  
 QY 3 HSILPRRAVVPVPLRMQ-----GPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPV 56  
 DB 744 HPSLPRSASTPQPIQOQSSIPPPPPPPHCEPT--MVHVEFTPTSTSSVPPPPPLP 801  
 QY 57 PVSVGRGPPD-----AHAVVN 73  
 DB 802 PISSGAPPPPPPPPPGIMHVAAS 825

## RESULT 10

A96826

T8K14.10 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: A96826

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasey, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malli, R.; Matzla

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: A96826

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-356 <STO>

A:Cross-references: GB:AE005173; NID:g4835761; PIDN:AAD30228.1; GSPDB:GN00141

C:Genetics:

A:Gene: TBK14.10

A:Map position: 1

#### Query Match

Best Local Similarity 18.2%; Score 76; DB 2; Length 356;  
 Pred. No. 2.9;

Matches 26; Conservative 7; Mismatches 30; Indels 12; Gaps 3;

QY 9 PAAPVPLRMQGPAPHVLSFLRPSMDVSAF-----YSLPLAPLSPTSPISV---VSY 60

DB 216 PGVPLGPPVSEPGSPPTGSPSPS---SGFLPPIVPPPMAPSPSVTPRTSAWCVAK 271

QY 61 GRGPPDAHVAVNLS 75

DB 272 PSVPDPPIQEAAMFA 286

#### RESULT 11

S48273

Probable transcription factor YBR108w - Yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein YBR0901

C:Species: *Saccharomyces cerevisiae*

C:Date: 01-Aug-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002

C:Accession: S48273; S45976; S44688

R:Manhaupt, G.; Stucka, R.; Ehme, S.; Vetter, I.; Feldmann, H.

Yeast 10.1363-1381.1994

A>Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.

A:Reference number: S48255; MUID:95208357; PMID:790426

A:Accession: S48273

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-848 <MAN>

A:Cross-references: EMBL:X78993; NID:g476045; PIDN:CA55611.1; PID:g476064

R:Feldmann, H.; Manhaupt, G.; Schwarzlose, C.; Vetter, I.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45927

A:Accession: S45976

A:Molecule type: DNA

A:Residues: 1-848 <FE2>

A:Cross-references: EMBL:Z35977; NID:g536378; PID:g536379; MIPS:YBR108w

C:Genetics:

A:Cross-references: SCD:S0000312

A:Map position: 2R

#### Query Match

Best Local Similarity 18.2%; Score 76; DB 2; Length 848;  
 Pred. No. 7.9;

Matches 23; Conservative 6; Mismatches 28; Indels 10; Gaps 3;

QY 13 PVLRLMQGPAPHV-----LSFLRPSMDVSAF---FYSPLAPLSPTSPV---ISPSVGR 62

DB 383 PVLVRMQGPAPHV-----LSFLRPSMDVSAF---FYSPLAPLSPTSPV---ISPSVGR 442

QY 63 GRPDDAH 69

DB 443 LPPEPTN 449

#### RESULT 12

D49600

genome-linked protein VPg - soybean dwarf virus

C:Species: soybean dwarf virus

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Sep-1999

C:Accession: D49600

R:Ratjen, J.P.; Karageorgos, L.E.; Habili, N.; Waterhouse, P.M.; Symons, R.H.

Virology 198, 671-679, 1994

A>Title: Soybean dwarf luteovirus contains the third variant genome type in the luteo

A:Reference number: A49600; MUID:94120742; PMID:8291248

A:Accession: D49600

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-189 <RAT>

A:Cross-references: GB:L24049; NID:g436017; PIDN:AAA17538.1; PID:g436021

C:Superfamily: potato leaf roll virus genome-linked protein

#### Query Match

Best Local Similarity 17.9%; Score 75; DB 2; Length 189;  
 Pred. No. 1.8;

Matches 19; Conservative 12; Mismatches 22; Indels 8; Gaps 2;

QY 2 THSLRPAAVPPVPL-----RMQGPAPHVLSFLRPSMDVSAF---FYSPLAPLSPTSPV 53

DB 58 THSCFQRTASMVVPREVLSGRLYONASLSLMEYRPTMINISRYSSSSRPLPPROY 117

QY 54 P 54

DB 118 P 118

#### RESULT 13

T48814

hypothetical protein 15E6.220 [Imported] - *Neurospora crassa*

C:Species: *Neurospora crassa*

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 05-May-2000

C:Accession: T48814

R:Schulte, U.; Algn, V.; Hohnsels, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu

submitted to the Protein Sequence Database, April 2000

A:Reference number: 224541

A:Accession: T48814

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1952 <SCH>

A:Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:15E6.220

A:Experimental source: cosmid contig 15E6; strain 74

C:Genetics:

A:Gene: NCSP:15E6.220

A:Map position: 2

A:introns: 281/3

#### Query Match

Best Local Similarity 17.9%; Score 75; DB 2; Length 1952;  
 Pred. No. 26;

Matches 24; Conservative 8; Mismatches 31; Indels 6; Gaps 3;

QY 2 THSLRPAAVPPVPLRMQGPAPHVLSFLRPSMDVSAF---FYSPLAPLSPTSPV---ISPV 58

DB 1781 TPQAVPRPSTSLTPRAQGPVSPAVS--GSGVAPASAOVAPAPVSSPTVPATVAPA 1838

QY 59 S-VGRGPPD 66

DB 1839 STVAAAPTP 1847

#### RESULT 14

T13029

beta-adaptin homolog FBL21.170 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 24-Nov-1999

C:Accession: T13029

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.;

submitted to the Protein Sequence Database, July 1999

A:Reference number: 217587

A:Accession: T13029

A:Molecule type: DNA

A:Residues: 1-894 <BEV>

A:Cross-references: EMBL:AL096882; GSPDB:GN00062; ATSP:FBL21.170

A:Experimental source: cultivar Columbia; BAC clone FBL21

C:Genetics:

A:Gene: ATSP:FBL21.170

A:Map position: 4

A:introns: 29/3; 95/3; 133/3; 217/2; 262/3; 322/3; 353/3; 424/2; 445/3; 505/3; 536/3;





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:25:15 ; Search time 2.85542 Seconds  
(without alignments)  
1147.512 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 418  
Sequence: 1 GTHSLLPRAVAVPLRMQP.....VGRGPDPAHVAVNLRYEG 79

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.5	20.0	888	SM6B_HUMAN	Q9h1t3 homo sapien
2	81.5	19.5	1257	PGCN_RAT	P53067 rattus norv
3	80	19.1	428	FXB2_MOUSE	Q64733 mus musculu
4	80	19.1	1840	RPB1_ARATH	P18616 arabidopsi
5	78	18.7	2319	NTC3_RAT	Q9r172 rattus norv
6	76	18.2	848	RYB8_YEAST	P38266 saccharomyc
7	74.5	17.8	1235	PER2_HUMAN	O15055 homo sapien
8	73.5	17.6	864	YCI8_HUMAN	Q9u1k2 homo sapien
9	73.5	17.6	864	SM6B_RAT	Q70141 rattus norv
10	73.5	17.6	1822	ZAP3_HUMAN	P49750 homo sapien
11	72	17.2	440	NK4R_HUMAN	P30098 homo sapien
12	71.5	17.1	828	SOX6_HUMAN	P35712 homo sapien
13	71.5	17.1	2441	CBP_MOUSE	P43481 mus musculu
14	70.5	16.9	291	YD53_STNY3	P42350 synchocyst
15	70.5	16.9	788	DPOL_HPBHE	P13346 heron hepat
16	70.5	16.9	1268	PGCN_MOUSE	P55066 mus musculu
17	70	16.7	2318	NTC3_MOUSE	Q61982 mus musculu
18	70	16.7	2321	NTC3_HUMAN	Q9um47 homo sapien
19	69.5	16.6	433	TEA3_CHICK	O90701 gallus gall
20	69	16.5	265	CDX1_HUMAN	P47902 homo sapien
21	68.5	16.4	886	SM6B_MOUSE	O54951 mus musculu
22	68	16.3	703	ML51_HUMAN	O15234 homo sapien
23	68	16.3	1206	FM14_MOUSE	Q05859 mus musculu
24	68	16.3	1468	FMN1_MOUSE	Q05860 mus musculu
25	68	16.3	1752	RPB1_SCHPO	P36594 schizosacch
26	68	16.3	1852	RPB1_CAEEL	P16356 caenorhabd
27	68	16.3	3298	PC16_HUMAN	Q96190 homo sapien
28	67.5	16.1	123	VST1_HEVBU	P29325 hepatitis e
29	67.5	16.1	123	VST1_HEVBY	O04612 hepatitis e
30	67.5	16.1	213	AMEX_BOVIN	P02817 bos taurus
31	67.5	16.1	506	ARSA_MOUSE	P50428 mus musculu
32	67.5	16.1	667	SIX5_MOUSE	P70178 mus musculu
33	67.5	16.1	827	SOX6_MOUSE	P40645 mus musculu

34	67	16.0	360	1	A2HS_RABIT	P80191 oryctolagus
35	67	16.0	817	1	VRP1_YEAST	P37370 saccharomyc
36	67	16.0	1433	1	Y310_HUMAN	O15027 homo sapien
37	66.5	15.9	259	1	MSP8_ETMAC	P09125 eimeria ace
38	66.5	15.9	283	1	EXTN_SORBI	P24152 sorghum bic
39	66.5	15.9	322	1	RX1_XENLA	O42201 xenopus lae
40	66.5	15.9	435	1	TEA3_HUMAN	O99594 homo sapien
41	66.5	15.9	1229	1	NI21_HUMAN	O97293 homo sapien
42	66	15.8	558	1	TF65_CHICK	P98152 gallus gall
43	66	15.8	736	1	ORP9_HUMAN	O96504 homo sapien
44	66	15.8	868	1	NRG2_RAT	O35569 rattus norv
45	66	15.8	1447	1	DCC_HUMAN	P43146 homo sapien

## ALIGNMENTS

RESULT 1  
ID SM6B\_HUMAN STANDARD: PRT; 888 AA.  
AC Q9H3T3; Q9NRK9;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Semaphorin 6B precursor (Semaphorin Z) (Sema Z).  
GN SEMA6B OR SEMA2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Brain;  
RA Kimura T., Ishida H.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RX MEDLINE-21248680; PubMed-11350127;  
RA Correa R.G., Sasahara R.M., Bengtson M.H., Katayama M.L.H.,  
RA Salim A.C.M., Brentani M.M., Sogayar M.C., de Souza S.J.,  
RA Simpson A.J.G.;  
\*Human semaphorin 6b.\*;  
Genomics 73:343-348(2001).  
CC -!- FUNCTION: MAY PLAY A ROLE IN BOTH PERIPHERAL AND CENTRAL NERVOUS  
CC -!- SYSTEM DEVELOPMENT (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- ALTERNATIVE PRODUCTS: At least 2 isoforms; 1 (shown here) and  
CC 2/6b.1; are produced by alternative splicing.  
CC -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AB022433; BAB20669.1; -;  
DR EMBL; AF216389; AAF87661.1; -;  
DR Genew; HGNC:10739; SEMA6B.  
DR InterPro: IPR003659; Plexin-like.  
DR InterPro: IPR001627; Sema.  
DR InterPro: IPR000737; Squash.  
DR Pfam; PF01403; Sema; 1.  
DR SMART; SM00423; PSI; 1.  
DR SMART; SM00286; PTI; 1.  
KW Signal. Transmembrane. Multigene family. Neurogenesis; Glycoprotein;  
KW Developmental protein; Alternative splicing.  
FT SIGNAL 1 25  
FT CHAIN 26 888 SEMAPHORIN 6B.  
FT DOMAIN 26 603 EXTRACELLULAR (POTENTIAL).

```

FT TRANSMEM 604 624 POTENTIAL.
FT DOMAIN 625 888 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 238 547 SEMA.
FT DOMAIN 661 674 POLY-GLY.
FT DOMAIN 750 753 POLY-LEU.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 471 517 LSVFLFEETFRPDRCGRGGEGTCGLLSLELDAASGGL
AAFPRC -> RVOCVGHACRVCHERSRSPORPGWLSRR
MGFORARGPORCLVY (IN ISOFORM 1).
FT VARSPPLIC 518 888 MISSING (IN ISOFORM 2).
FT CONFLICT 30 30 D -> E (IN REF. 2).
SQ SEQUENCE 888 AA: 95270 MM: 6FFB4D6828C70CB CRC64:

Query Match 20.0%; Score 83.5; DB 1; Length 888;
Best Local Similarity 35.4%; Pred. No. 0.8;
Matches 28; Conservative 7; Mismatches 33; Indels 11; Gaps 5;

OY 1 GTH-----SLPRPAVPLRMQGPAPHVLSFLRP-SMD---LVSAFYSLPLAPLSPT 51
DB 698 GPHDLDSGLLPTEQHPRLPQKRLPPT-HPHPHALGPRAMDHGHPLLPASASSLLAPLA 756
OY 52 SVIPSPVSGRGPDPDAH 70
DB 757 RAPEOPPAGE-PTPDGRL 774

RESULT 2
PCGN_RAT STANDARD: PRT: 1257 AA.
AC P53067;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Annotation update)
DE Neurocan core protein precursor (245 kDa early postnatal core
glycoprotein) [Contains: 150 kDa adult core glycoprotein].
GN CSPG3 OR NCAN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OY NCBL_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=92406907; PubMed=1326557;
RA Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;
RT "Cloning and primary structure of neurocan, a developmentally
regulated, aggregating chondroitin sulfate proteoglycan of brain.";
RL J. Biol. Chem. 267:19536-19547(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=94230574; PubMed=7513709;
RA Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,
Margolis R.U., Grumet M.;
RT "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
neural cell adhesion molecules Ng-CAM/11/NILE and N-CAM, and inhibits
neural cell adhesion and neurite outgrowth.";
RL J. Cell Biol. 125:669-680(1994).
CC -1- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING
DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NG-CAM
AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC
ACID.
CC -1- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED
IN KIDNEY, LUNG, LIVER AND MUSCLE.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -1- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL

```

```

CC CC BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M97161; AAC37679.1; -.
CC HSSP: P00740; IEDM.
CC DR Interpro: IPR000152; Asx_hydroxyl.
CC DR Interpro: IPR000561; EGF-like.
CC DR Interpro: IPR000742; EGF-like.
CC DR Interpro: IPR001881; EGF_Ca.
CC DR Interpro: IPR003599; Ig.
CC DR Interpro: IPR003006; Ig_MHC.
CC DR Interpro: IPR001304; Lectin_C.
CC DR Interpro: IPR000538; Link.
CC DR Interpro: IPR000436; Sushl_SCR_CCP.
CC DR Pfam: PF00008; EGF_2.
CC DR Pfam: PF00047; Ig_1.
CC DR Pfam: PF00059; Lectin_C_1.
CC DR Pfam: PF00084; sushl_1.
CC DR Pfam: PF00193; Xlink_2.
CC DR ProDom: PD000918; Link_2.
CC DR SMART: SM00032; CCP_1.
CC DR SMART: SM00034; CLECT_1.
CC DR SMART: SM00179; EGF_CA_1.
CC DR SMART: SM00001; EGF_Like_1.
CC DR SMART: SM00409; Ig_1.
CC DR SMART: SM00445; Link_2.
CC DR PROSITE: PS00010; ASX_HYDROXYL_1.
CC DR PROSITE: PS00022; EGF_1; 3.
CC DR PROSITE: PS0186; EGF_2; 1.
CC DR PROSITE: PS0187; EGF_CA_1.
CC DR PROSITE: PS01241; LINK_2.
CC DR PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
CC DR PROSITE: PS50041; C-TYPE-LECTIN_2; 1.
CC KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushl; Signal.
FT FT SIGNAL 1 22
FT CHAIN 23 1257
FT DOMAIN 639 1257
FT DOMAIN 51 146
FT DOMAIN 176 253
FT DOMAIN 274 355
FT DOMAIN 949 985
FT DOMAIN 987 1023
FT DOMAIN 1025 1154
FT DOMAIN 1155 1213
FT DISULFID 38 139
FT DISULFID 181 252
FT DISULFID 205 226
FT DISULFID 279 354
FT DISULFID 303 324
FT DISULFID 953 964
FT DISULFID 958 973
FT DISULFID 975 984
FT DISULFID 1029 1040
FT DISULFID 1057 1149
FT DISULFID 1125 1141
FT DISULFID 1156 1199
FT DISULFID 1185 1212
FT CARBOHYD 121 121
FT CARBOHYD 339 339

```

```

FT CAROHD 737 737 N-LINKED (GLCNC. . .) (POTENTIAL).
FT CAROHD 944 944 O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
FT CAROHD 967 967 N-LINKED (GLCNC. . .) (POTENTIAL).
FT CAROHD 1164 1164 N-LINKED (GLCNC. . .) (POTENTIAL).
SQ SEQUENCE 1257 AA; 135544 MW; 992B33DCFA19EEEB CRC64;

Query Match
Best Local Similarity 19.5%; Score 81.5; DB 1; Length 1257;
Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

QY 4 SILPRPAVPVPLRMQPG---PAHPVLSFLR-----PSMDLVSAFYSLPLAPLS--PT 51
DB 610 SSIPSEALSAVSLQASPGSGSPDFIVAMLRAPKMLPLPHSTLVNVPSPILSPASPLPS 669
QY 52 SVP-----ISPVSVGRGPPDP 67
DB 670 SYPEEQAVNPVSFG--AEDE 688

RESULT 3
FXB2_MOUSE STANDARD: PRT: 428 AA.
AC 064733:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forhead box protein B2 (Transcription factor FKX-4).
GN FOXB2 OR FKX4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97014266; PubMed=8861101;
RA Kaestner K.H., Schuetz G., Monaghan A.P.;
RT "Expression of the winged helix genes fkh-4 and fkh-5 defines domains
RT in the central nervous system.";
RL Mech. Dev. 55:221-230(1996).
RN [2]
RP SEQUENCE OF 4-114 FROM N.A.
RC STRAIN=129;
RX MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,
RA Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forhead gene family are developmentally
RT regulated.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
RN [3]
RP SUBCELLULAR LOCATION: Nuclear.
CC -I- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -I- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X92591; CA63335.1; -
DR EMBL: X71942; CA50744.1; -
DR HSSP: O63345; 2HPH.
DR TRANSFAC: T02442; -
DR MGD: MGI:1347468; Foxb2.
DR InterPro: IPR001766; TF_Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR PRODOM: PD000425; TF_Fork_head; 1.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.

```

```

KW DNA-binding; Nuclear protein; Transcription regulation.
FT DNA_BIND 12 103 FORK-HEAD.
FT DOMAIN 139 153 POLY-HIS.
FT DOMAIN 156 162 POLY-HIS.
FT DOMAIN 163 172 POLY-PRO.
FT DOMAIN 217 231 POLY-ALA.
FT DOMAIN 249 258 POLY-ALA.
FT DOMAIN 321 330 POLY-ALA.
FT DOMAIN 396 399 POLY-ALA.
SQ SEQUENCE 428 AA; 45170 MW; DB8ABEFD1E94AB10 CRC64;

Query Match
Best Local Similarity 19.1%; Score 80; DB 1; Length 428;
Matches 23; Conservative 10; Mismatches 25; Indels 8; Gaps 2;

QY 3 HSLLPRPAVPVPLRMQPGPAHPVLSFLRPSMDLVSAFYSLP-----LAPLSPTSVIS 56
DB 350 HSANQSEPAVPVPIK--PTPALPVTTLTPALSVPTASQQLPAPSTVCMAASPTAPILLE 407
QY 57 PVSGR 62
DB 408 PTAAGR 413

RESULT 4
RBP1_ARATH STANDARD: PRT: 1840 AA.
ID RBP1_ARATH
AC P18616; P31635; Q9S2S8;
DT 01-NOV-1990 (Rel. 16, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6).
GN RPB205 OR RPII OR RPB1 OR AT4G35800 OR F4B14.70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=91080867; PubMed=2259344;
RA Nawrath C., Scheil J., Koncz C.;
RT "Homologous domains of the largest subunit of eucaryotic RNA
RT polymerase II are conserved in plants.";
RL Mol. Gen. Genet. 223:65-75(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=91355869; PubMed=2103447;
RA Dietrich M.A., Prenger J.P., Guilfoyle T.J.;
RT "Analysis of the genes encoding the largest subunit of RNA polymerase
RT II in Arabidopsis and soybean.";
RL Plant Mol. Biol. 15:207-223(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083486; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entlan K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
RA Kreisel M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoer E.,
RA Weltzienegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Stavelen M., Dirkse W.,
RA Woolman P., Klein Lankhorst R., Rose W., Hauf J., Koetter P.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gieten J., Villarroel R., De Clercq R.,

```





```
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- TISSUE SPECIFICITY: Expressed in postnatal central nervous system
CC (CNS) germinal zones and, in early postnatal life, within
CC numerous cells throughout the CNS. It is more highly localized
CC to ventricular germinal zones.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-cis network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF164486; AAD46653.2; -.
CC DR HSSP: P00740; 1EDM.
CC DR InterPro: IPR002110; ANK.
CC DR InterPro: IPR000152; Asx_hydroxyl.
CC DR InterPro: IPR000561; EGF-like.
CC DR InterPro: IPR000742; EGF-2.
CC DR InterPro: IPR001881; EGF-Ca.
CC DR InterPro: IPR001438; EGF-11.
CC DR InterPro: IPR002049; Laminin_EGF.
CC DR InterPro: IPR000800; Notch.
CC DR Pfam: PR00008; EGF; 34.
CC DR Pfam: PF00023; ank; 6.
CC DR Pfam: PF00066; notch; 3.
CC DR PRINTS: PR00010; EGFBL00D.
CC DR PRINTS: PR00011; EGF_LAMININ.
CC DR PRINTS: PR01452; NOTCH.
CC DR SMART: SM00248; ANK; 5.
CC DR SMART: SM00179; EGF_CA; 20.
CC DR SMART: SM00001; EGF_Like; 12.
CC DR PROSITE: PS50088; ANK_REPEAT; 4.
CC DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
CC DR PROSITE: PS00010; ASX_HYDROXYL; 18.
CC DR PROSITE: PS00022; EGF-1; 33.
CC DR PROSITE: PS01186; EGF-2; 26.
CC DR PROSITE: PS01187; EGF_CA; 17.
CC Receptor: Transcription regulation: Activator; Differentiation;
CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
CC KW SIGNAL.
CC FT CHAIN 1 2319 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
CC FT CHAIN 1631 2319 NOTCH EXTRACELLULAR TRUNCATION (BY
CC FT CHAIN 1631 2319 SIMILARITY).
CC FT CHAIN 1664 2319 NOTCH INTRACELLULAR DOMAIN (BY
CC FT CHAIN 1664 2319 SIMILARITY).
CC FT DOMAIN 41 1645 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 1646 1666 POTENTIAL.
CC FT DOMAIN 1667 2319 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 41 79 EGF-LIKE 1.
CC FT DOMAIN 80 120 EGF-LIKE 2.
CC FT DOMAIN 121 158 EGF-LIKE 3.
CC FT DOMAIN 160 197 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 199 236 EGF-LIKE 5.
FT DOMAIN 238 274 CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 276 314 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 316 352 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 353 391 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 393 431 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 433 469 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 471 507 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 509 545 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 547 582 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 584 620 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 622 657 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 659 695 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 697 732 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 736 772 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 773 810 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 812 849 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 851 887 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 889 924 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 926 962 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 964 1000 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1002 1036 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1038 1084 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1086 1122 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1124 1160 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1162 1205 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1207 1246 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1248 1289 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1291 1327 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1329 1375 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1377 1420 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 1425 1461 LIN/NOTCH 1.
FT REPEAT 1463 1503 LIN/NOTCH 2.
FT REPEAT 1505 1542 LIN/NOTCH 3.
FT REPEAT 1544 1573 ANK 1.
FT REPEAT 1575 1612 ANK 2.
FT REPEAT 1614 1651 ANK 3.
FT REPEAT 1653 1690 ANK 4.
FT REPEAT 1692 1729 ANK 5.
FT SITE 1731 1774 CLEAVAGE BY FURIN-LIKE PROTEASE (BY
FT DISULFID 44 56 SIMILARITY).
FT DISULFID 50 67 BY SIMILARITY.
FT DISULFID 69 78 BY SIMILARITY.
FT DISULFID 84 95 BY SIMILARITY.
FT DISULFID 89 108 BY SIMILARITY.
FT DISULFID 110 119 BY SIMILARITY.
FT DISULFID 125 136 BY SIMILARITY.
FT DISULFID 130 146 BY SIMILARITY.
FT DISULFID 148 157 BY SIMILARITY.
FT DISULFID 164 176 BY SIMILARITY.
FT DISULFID 170 185 BY SIMILARITY.
FT DISULFID 187 196 BY SIMILARITY.
FT DISULFID 203 214 BY SIMILARITY.
FT DISULFID 208 224 BY SIMILARITY.
FT DISULFID 226 235 BY SIMILARITY.
FT DISULFID 242 253 BY SIMILARITY.
FT DISULFID 247 262 BY SIMILARITY.
FT DISULFID 264 273 BY SIMILARITY.
FT DISULFID 280 293 BY SIMILARITY.
FT DISULFID 287 302 BY SIMILARITY.
FT DISULFID 304 313 BY SIMILARITY.
FT DISULFID 330 331 BY SIMILARITY.
FT DISULFID 335 340 BY SIMILARITY.
FT DISULFID 342 351 BY SIMILARITY.
FT DISULFID 357 368 BY SIMILARITY.
FT DISULFID 362 379 BY SIMILARITY.
FT DISULFID 381 390 BY SIMILARITY.
FT DISULFID 397 410 BY SIMILARITY.
FT DISULFID 404 419 BY SIMILARITY.
FT DISULFID 421 430 BY SIMILARITY.
FT DISULFID 437 448 BY SIMILARITY.
FT DISULFID 442 457 BY SIMILARITY.
FT DISULFID 459 468 BY SIMILARITY.
```

```

FT DISULFID 475 486 BY SIMILARITY.
FT DISULFID 480 495 BY SIMILARITY.
FT DISULFID 497 506 BY SIMILARITY.
FT DISULFID 513 524 BY SIMILARITY.
FT DISULFID 518 533 BY SIMILARITY.
FT DISULFID 535 544 BY SIMILARITY.
FT DISULFID 551 561 BY SIMILARITY.
FT DISULFID 556 570 BY SIMILARITY.
FT DISULFID 572 581 BY SIMILARITY.

Query Match 18.7%; Score 78; DB 1; Length 2319;
Best Local Similarity 39.3%; Pred. No. 7.5;
Matches 24; Conservative 2; Mismatches 23; Indels 12; Gaps 1;

QY 6 LPRPAVPVPLMOPGAPVLSFLRPSMDLVSASFSLPLAPLSPTSPVSGRGPD 65
DB 2163 LNPVAVPLDMARLPPAPPSFL-----LPLAPGSQLNPATPVSPHERPP 2210

QY 66 P 66
DB 2211 P 2211

RESULT 6
YB8_YEAST STANDARD: PRT: 848 AA.
ID YB8_YEAST
AC P38266;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1994 (Rel. 37, Last annotation update)
DE Hypothetical 92.8 kDa protein in PHO88-CMD1 intergenic region.
GN YBR108W OR YBR0901.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=9508357; PubMed=7900426;
RA Mannhaupt G., Stucka R., Ehmlé S., Vetter I., Feldmann H.;
RT Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL Yeast 10:1363-1381(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X78993; CAA55611.1; -
DR EMBL: Z35977; CAA85063.1; -
DR PIR: S44688; S44688.
DR SGD: S0000312; YBR108W.
KW Hypothetical protein.
SQ SEQUENCE 848 AA; 92762 MW; F33D371369FBAF97 CRC64;

Query Match 18.2%; Score 76; DB 1; Length 848;
Best Local Similarity 34.3%; Pred. No. 4.1;
Matches 23; Conservative 6; Mismatches 28; Indels 10; Gaps 3;

QY 13 PVLPMOPGAPV-----LSFLRPSMDLVS---FYSFLAPLSPTSPVSGR 62
DB 383 PVPVRRQPPQPPMOGNITPIEPSLDTGSTRPHEVTPFPDPAAPKPKIDIPVDVSS 442

QY 63 GPDPDAH 69
DB 443 LPPPTH 449

RESULT 7

```

```

PER2_HUMAN STANDARD: PRT: 1255 AA.
ID PER2_HUMAN
AC O15055;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Period circadian protein 2.
GN PER2 OR KIAA0347.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC TISSUE=Brain;
RA Nagase T., Ishikawa K.-I., Seki N., Nakajima D., Ohira M.,
RA Miyajima N., Kotani H., Nomura N., Ohara O.;
RL Submitted (Dec-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP TISSUE EXPRESSION.
RX MEDLINE=98087121; PubMed=9427249;
RA Sheerman L.P., Zylka M.J., Weaver D.R., Kolakowski L.F. Jr.,
RA Reppert S.M.;
RT "Two period homologs: circadian expression and photic regulation in
RT the suprachiasmatic nuclei.";
RL Neuron 19:1261-1269(1997).
CC -----
CC -1- FUNCTION: CIRCADIAN REGULATOR THAT MAY ACT AS A TRANSCRIPTION
CC FACTOR. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN TRANSCRIPTIONAL
CC TRANSCRIPTIONAL INHIBITION. EXPRESSION OSCILLATES WITH A 24 HR
CC RHYTHM IN THE SUPRACHIASMATIC NUCLEUS (SCN) AND THE WHOLE EYES.
CC OSCILLATIONS ARE MAINTAINED UNDER CONSTANT DARKNESS AND ARE
CC RESPONSIVE TO CHANGES OF THE LIGHT/DARK CYCLES. THERE IS A 4 HOUR
CC TIME DELAY BETWEEN PER1 AND PER2 OSCILLATIONS. THE EXPRESSION
CC RHYTHMS APPEAR TO ORIGINATE FROM RETINA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. FOUND IN HEART, BRAIN,
CC PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS.
CC -1- INDUCTION: BY LIGHT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB002345; BAA20804.2; ALT_INIT.
DR Genew: HGNC:8846; PER2.
DR MIM: 603426;
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS_domain.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00112; PAS; 1.
KW Transcription regulation; Nuclear protein; Repeat; Biological rhythms.
FT DOMAIN 109 146 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 182 248 PAS 1.

```

```

FT DOMAIN 319 389 PAS.2.
FT DOMAIN 398 438 PAC.
FT DOMAIN 510 513 POLY-ARG.
FT DOMAIN 789 806 NUCLEAR LOCALIZATION SIGNAL.
FT DOMAIN 842 979 PRO-RICH.
SQ SEQUENCE 1255 AA; 136579 MW; 2AEF2C6BD46CB80 CRC64;

Query Match
Best Local Similarity 17.8%; Score 74.5; DB 1; Length 1255;
Matches 30; Conservative 10; Mismatches 26; Indels 51; Gaps 5;

OY 1 GTHSLPRF-----AAYVPLRMQ-----PGPAHVLSTLRPSMDLVSAFSLPL 45
DB 863 GYVAAPPAHPHASFYVPAVPDLQHOFVAVQPPFPAPLAPVMAFMLPSYSPSGRTNLPQ 922

OY 46 A-PLSP-----TSVP-----ISPVSVGRGPP 66
DB 923 AEFPSOPFPSPHPTLTSEMASASQPEFPSPRTSIPROPACAPATRAVTPPSAMGRASPP 979

RESULT 8
YC18_HUMAN STANDARD; PRT; 864 AA.
AC 09ULK2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA1218 (Fragment).
GN KIAA1218.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:337-345(1999).
CC -1- SIMILARITY: BELONGS TO THE ATAXIN 7 FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL AB033044; BAA86532.1; -
KW Hypothetical protein.
FT NON_TER 1
FT DOMAIN 223 226 POLY-THR.
FT DOMAIN 552 664 POLY-SER.
FT DOMAIN 786 790 POLY-SER.
SQ SEQUENCE 864 AA; 92138 MW; DESC8130E48DA23 CRC64;

Query Match
Best Local Similarity 17.6%; Score 73.5; DB 1; Length 864;
Matches 33; Conservative 4; Mismatches 21; Indels 35; Gaps 7;

OY 4 SLTPRPA--VPVLR-MQGPAPHVLSTLRPSMDLVSAFSLPLASP-----T 51
DB 517 SPLSPAPAHITTPVPAVYLQ-----FSNP-----SAVY-LPSAPISRLTSYIMT 562

OY 52 SVPISPVSVGRGPPDA-----HVAVNLS 75
DB 563 SAMLSNAAFVTSPPDSALMSHTTAPFPHVAATLS 595

```

```

RESULT 9
SM6B_RAT STANDARD; PRT; 887 AA.
AC 070141;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Semaphorin 6B precursor (Semaphorin Z) (Sema Z).
GN SEMA6B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=Mistat; TISSUE=Brain;
RX MEDLINE=98087397; PubMed=9427525;
RA Kikuchi K., Ishida H., Kimura T.;
RT "Molecular cloning of a novel member of semaphorin family genes,
RT semaphorin Z."
RL Brain Res. Mol. Brain Res. 51:229-237(1997).
CC -1- FUNCTION: MAY PLAY A ROLE IN BOTH PERIPHERAL AND CENTRAL NERVOUS
CC SYSTEM DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: DETECTED IN THE FIRST BRANCHIAL ARCH OF
CC EMBRYONIC DAY 11 (E11) EMBRYO. AND SUBSEQUENTLY IN THE MYOTOMES
CC AND THE DORSAL ROOT GANGLIA IN DEVELOPING SOMITES FROM E11.5
CC THROUGH E13.5, BUT NOT IN THE BRAIN. HOWEVER, AT E15, 18, 21 AND
CC P0, SEMA2 WAS HIGHLY EXPRESSED IN THE BRAIN.
CC -1- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -1- SIMILARITY: CONAINS 1 SEMA DOMAIN.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL AB000776; BAA25687.1; -
DR InterPro: IPR003659; Plexin-like.
DR InterPro: IPR001627; Sema.
DR Pfam: PF01403; Sema; 1.
DR SMART: SM00423; PST; 1.
KW Signal; Transmembrane; Multigene family; Neurogenesis; Glycoprotein;
KW developmental protein.
FT SIGNAL 1 26
FT CHAIN 27 887 SEMAPHORIN 6B.
FT DOMAIN 27 605 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 606 626 POTENTIAL.
FT DOMAIN 627 887 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 239 549 SEMA.
FT CARBOHYD 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 887 AA; 95752 MW; 09543F3F202CD301 CRC64;

Query Match
Best Local Similarity 17.6%; Score 73.5; DB 1; Length 887;
Matches 27; Conservative 6; Mismatches 31; Indels 11; Gaps 5;

OY 1 GTH-----SLTPPAVVPVLRMQGPAPHVLSTLRP-SMD-----LVSAFYSLLPLASPT 51
DB 699 GPHDLSGLPLPPEQTPPLPKRLP-TTHPHAHALPBRANDSHALLSASASTSLLLAHT 757

OY 52 SVPISPVSVGRGPP 65
DB 563 SAMLSNAAFVTSPPDSALMSHTTAPFPHVAATLS 595

```

```

Db      758 RAEPQPVPTESGPE 772

RESULT 10
ZAP3_HUMAN      STANDARD:      PRT: 1822 AA.
ID      ZAP3_HUMAN
AC      P49750; P49752; Q9P1V7;
DT      01-OCT-1996 (Rel. 34, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Nuclear protein ZAP3 (ZAP113).
GN      ZAP3.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Rowen L., Madan A., Qin S., Abbasi N., Baradaran L., Birditt B.,
RA      Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,
RA      Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.;
RT      "Sequencing of human chromosome 14q24.3 region.";
RL      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=95319502; PubMed=7596406;
RA      Sherrington R., Rogeev E.I., Liang Y., Rogeeva E.A., Levesque G.,
RA      Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
RA      Foinin J., F., Bruni A.C., Montesi M.P., Sorbi S., Ralnero I.,
RA      Plinsau P., Nee L., Chumakov I., Pollen D., Brookes A.,
RA      Sansau P., Polinsky R.J., Masco W., da Silva H.A.R., Haines J.L.,
RA      Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
RA      Rommens J.M., St George-Hyslop P.H.;
RT      "Cloning of a gene bearing missense mutations in early-onset familial
RT      Alzheimer's disease.";
RL      Nature 375:754-760(1995).
CC      -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC      -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC      frameshift in position 1661.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; AC007956; AAF61275.1; -
DR      EMBL; L40403; AAC42008.1; ALT_FRAME.
DR      EMBL; L40400; AAC42006.1; -
KW      Nuclear protein.
FT      DOMAIN 15 205 PRO-RICH.
FT      DOMAIN 382 430 GLN-RICH.
FT      DOMAIN 807 1209 ARG-RICH.
FT      DOMAIN 1488 1577 ARG-RICH.
FT      DOMAIN 621 621 P -> S (IN REF. 2).
FT      CONFLICT 1404 1404 T -> I (IN REF. 2).
FT      CONFLICT 1821 1821 K -> E (IN REF. 2).
SQ      SEQUENCE 1822 AA; 204947 MW; 8E6CB83FE540C7D2 CRC64;

Query Match      17.6%; Score 73.5; DB 1; Length 1822;
Best Local Similarity 35.0%; Pred. No. 16;
Matches 21; Conservative 7; Mismatches 23; Indels 9; Gaps 2;
OY      7 PPAAPVPLRMQPGA--HPVLSELRSMWLVSAFYSLPLAPISPTSPISPVSGRG 63
DB      471 PPAPLLPVGSGAPPTTYHPPLQSGAPSGEQVNSK-----APLSKSLAPYSSSSDQG 524

RESULT 11
NK4R_HUMAN

```

```

ID      NK4R_HUMAN      STANDARD:      PRT: 440 AA.
AC      P30098;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Neuromedin K receptor (NK4R) (Neurokinin B receptor) (NK-4 receptor)
DE      (NK-4R) (K1R) (Neurokinin 4 receptor) (NK4).
GN      TACR3L OR TAC3RL OR TCAR.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Placenta;
RX      MEDLINE=92237319; PubMed=1315051;
RA      Xie G.-X., Miyajima A., Goldstein A.;
RT      "Expression cloning of cDNA encoding a seven-helix receptor from
RT      human placenta with affinity for opioid ligands.";
RL      Proc. Natl. Acad. Sci. U.S.A. 89:4124-4128(1992).
RN      [2]
RP      CHARACTERIZATION.
RX      MEDLINE=97103087; PubMed=8947459;
RA      Donaldson L.F., Haskell C.A., Hanley M.R.;
RT      "Functional characterization by heterologous expression of a novel
RT      cloned tachykinin peptide receptor.";
RL      Biochem. J. 320:1-5(1996).
CC      -1- FUNCTION: THIS IS A RECEPTOR FOR THE TACHYKININ NEUROPEPTIDE
CC      NEUROMEDIN K (NEUROKININ B). THE RANK ORDER OF AFFINITY OF THIS
CC      RECEPTOR TO TACHYKININS IS: NEUROMEDIN K > SUBSTANCE K > SUBSTANCE
CC      P.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC      -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE, LIVER,
CC      LUNG AND HEART. LOW LEVELS IN PANCREAS.
CC      -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC      HIGHEST TO OTHER TACHYKININS RECEPTORS.
CC      -1- CAUTION: Was originally (Ref.1) thought to be a kappa-type opioid
CC      receptor.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; M84605; AAA36395.1; -
DR      PIR; A44081; A44081.
DR      HSSP; P02699; 1E88.
DR      InterPro; IPR000276; GPCR_Rhodpsn.
DR      Pfam; PF00001; 7tm.1; 1.
DR      PRINTS; PR00237; GPCRHHODPSN.
DR      PROSITE; PS00237; G-PROTEIN_RECPEP_FL_1; 1.
DR      PROSITE; PS50262; G-PROTEIN_RECPEP_FL_2; 1.
KW      G-protein coupled receptor; Transmembrane; Glycoprotein; Lipoprotein;
KW      Palmitate.
FT      DOMAIN 1 59 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 60 82 1 (POTENTIAL).
FT      DOMAIN 83 92 CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 93 114 2 (POTENTIAL).
FT      DOMAIN 115 134 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 135 156 3 (POTENTIAL).
FT      DOMAIN 157 176 CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 177 197 4 (POTENTIAL).
FT      DOMAIN 198 220 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 221 245 5 (POTENTIAL).
FT      DOMAIN 246 274 CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 275 296 6 (POTENTIAL).
FT      DOMAIN 297 309 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 310 334 7 (POTENTIAL).
FT      DOMAIN 335 440 CYTOPLASMIC (POTENTIAL).
FT      DISULFD 133 208 BY SIMILARITY.

```



```

CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S66385; AAB28651.1; -
CC TRANSFAC: T01318; -
CC MGD: MGI:1098280; Crebbp.
CC InterPro: IPR001487; Bromodomain.
CC InterPro: IPR003101; KIX.
CC InterPro: IPR000197; TAZ_finger.
CC InterPro: IPR000433; Zn1_Z2.
CC Pfam: PF00439; bromodomain; 1.
CC Pfam: PF00569; Z2; 1.
CC Pfam: PF02135; zf-TAZ; 2.
CC Pfam: PF02172; KIX; 1.
CC PRINTS: PR00503; BROMODOMAIN.
CC SMART: SM00297; BROMO; 1.
CC SMART: SM00291; ZnF_Z2; 1.
CC PROSITE: PS00633; BROMODOMAIN_1; 1.
CC PROSITE: PS00014; BROMODOMAIN_2; 1.
CC PROSITE: PS01357; ZF_Z2_1; 1.
CC PROSITE: PS0135; ZF_Z2_2; 1.
CC Transcription regulation; Nuclear protein; Activator; Bromodomain;
CC Zinc-finger.
KW
FT DOMAIN 1104 1176 BROMODOMAIN.
FT ZN_FING 1702 1745 Z2-TYPE.
FT DOMAIN 1062 1065 POLY-GLU.
FT DOMAIN 1556 1563 POLY-GLU.
FT DOMAIN 1944 1949 POLY-PRO.
FT DOMAIN 1968 1971 POLY-GLN.
FT DOMAIN 2082 2086 POLY-GLN.
FT DOMAIN 2200 2216 POLY-GLN.
FT DOMAIN 2296 2299 POLY-GLN.
SO SEQUENCE 2441 AA; 265474 MW; 0ABBD28C3112F419 CRC64;

Query Match 17.1%; Score 71.5; DB 1; Length 2441;
Best Local Similarity 36.5%; Pred. No. 34;
Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

QY 4 SLPRPAAVVPLRMQGPAPHVLSFLRPSWDLVSATYSLPLAPLSTSPISPVSGRG 63
DB 843 SQLPCRPVPTOSPLHPPPPASTAGM--PSLQHPATGMPQPAPAPTO--PSTPVSSGQT 899

QY 64 PDP 66
DB 900 PTP 902

RESULT 14
YD53_SYNY3 STANDARD; PRT; 291 AA.
AC P42350;
ID YD53_SYNY3
OX NCBI_TaxID=1148;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=93222488; PubMed=8467083;
RA Malakhov M.P., Wada H., Los D.A., Sakamoto T., Murata N.;
RT "Structure of a cyanobacterial gene encoding the 50S ribosomal
protein L9.";
```

```

RL Plant Mol. Biol. 21:913-918(1993).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyura M., Saito S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RT DNA Res. 3:109-136(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10716; BAA38818.1; -
CC EMBL: D90912; BAA18174.1; -
CC Hypothetical protein; Complete proteome.
KW SEQUENCE 291 AA; 31283 MW; 5AB7E3DD03C36390 CRC64;

Query Match 16.9%; Score 70.5; DB 1; Length 291;
Best Local Similarity 28.8%; Pred. No. 4.5;
Matches 23; Conservative 16; Mismatches 30; Indels 11; Gaps 5;

QY 4 SLPRPAAVVPLRMQGPAPHVLSFLRPSWDLVSATYSLP--LAPLSPT-----SVP 54
DB 145 ALAPEPTLPPAPISPPSP--DPLSLSEPTPPPPAMVNSFNQPEESAPIDSELDQDFARP 203

QY 55 ISPVSYGRGD-PDAHVAVN 73
DB 204 ELPLAVEARKDPSPEPDMAVS 223

RESULT 15
DPOL_HPBHE
ID DPOL_HPBHE STANDARD; PRT; 788 AA.
AC P13846;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE P protein (includes: DNA-directed DNA polymerase (EC 2.7.7.7); RNA-
DE directed DNA polymerase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)).
GN
OS Heron hepatitis b virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=28300;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88333160; PubMed=3418788;
RA Sprengel R., Kallea E.F., Will H.;
RT "Isolation and characterization of a hepatitis B virus endemic in
RT herons.";
RL J. Virol. 62:3832-3839(1988).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + (DNA)(N).
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

DR EMBL: M22056; AAA45738.1; -.  
DR PIR: A30082; JDTLHH.  
DR InterPro: IPR001462; DNApol\_viral\_C.  
DR InterPro: IPR000201; DNApol\_viral\_N.  
DR InterPro: IPR000477; RTase.  
DR Pfam: PF00078; Vtl; 1.  
DR Pfam: PF00242; DNA\_pol\_viral\_N; 1.  
DR Pfam: PF00036; DNA\_pol\_viral\_C; 1.  
DR ProDom: PD000814; DNApol\_viral\_C; 1.  
KW Transferrase: RNA-directed DNA polymerase: DNA-directed DNA polymerase.  
KW Hydrolyase: Nuclease: Endonuclease; DNA replication; DNA-binding.  
SQ SEQUENCE 788 AA; 90070 MW; FB44F36F75EAD644 CRC64;

Query Match	16.9%	Score 70.5;	DB 1;	Length 788;
Best Local Similarity	40.0%;	Pred. No. 13;		
Matches 18; Conservative	4;	Mismatches 12;	Indels 11;	Gaps 1;

```

QY      31 RPSNDLSAIFYSLPLAPLSPTSVPI-----PVSVGRGP 64
          | | | | | | | | | | : : | | |
Db      444 RISLDSQAFYHLPPLAPASSRLAVSDGQVYVYFRKAPMGVLSP 488

```

Search completed: March 4, 2003, 12:34:57  
Job time : 4.85542 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:31:26 ; Search time 8.7249 seconds  
(without alignments)  
1865.663 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 418

Sequence: 1 GTHSLPRPAAVVPLRMQP.....VGRGPDPAHVAVNLSTREYEG 79

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched:

671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTRMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	406	97.1	419	4 Q9UK79	Q9UK79 homo sapien
2	99.5	23.8	309	12 Q8U2B4	Q8U2B4 grapevine f
3	85	20.3	327	11 Q9JUK6	Q9JUK6 mus musculu
4	84.5	20.2	434	5 Q9G015	Q9G015 theileria p
5	84	20.1	995	11 Q35615	Q35615 mus musculu
6	82.5	19.7	316	10 Q8R2Y1	Q8R2Y1 oryza sativ
7	82	19.6	122	12 Q91115	Q91115 hepatitis e
8	82	19.6	122	12 Q8V730	Q8V730 swine hepat
9	81.5	19.5	816	11 Q70474	Q70474 rattus norv
10	81	19.4	200	10 Q82761	Q82761 arabidopsis
11	81	19.4	1006	10 Q9LMQ1	Q9LMQ1 arabidopsis
12	78.5	18.8	941	12 Q9IMV0	Q9IMV0 cercopithec
13	78	18.7	122	12 Q9YLR0	Q9YLR0 hepatitis e
14	78	18.7	763	2 Q9XDH2	Q9XDH2 mycobacteri
15	77.5	18.5	487	5 Q9V197	Q9V197 drosophila
16	77.5	18.5	1212	16 Q9L1C8	Q9L1C8 streptomyces

17	77	18.4	122	12 Q36612	Q36612 swine hepat
18	77	18.4	503	5 Q17585	Q17585 caenorhabd
19	77	18.4	690	10 Q94D41	Q94D41 oryza sativ
20	77	18.4	1044	4 Q8W4X9	Q8W4X9 homo sapien
21	76.5	18.3	1110	5 Q09493	Q09493 caenorhabd
22	76	18.2	309	3 Q9P6R1	Q9P6R1 schizosacch
23	76	18.2	356	10 Q9SAK1	Q9SAK1 arabidopsis
24	76	18.2	1044	4 Q9NXC6	Q9NXC6 homo sapien
25	75	17.9	189	12 Q87034	Q87034 soybean dwa
26	75	17.9	189	12 Q91Q07	Q91Q07 soybean dwa
27	75	17.9	189	12 Q08404	Q08404 soybean dwa
28	75	17.9	295	16 Q9KXQ7	Q9KXQ7 streptomyce
29	75	17.9	813	10 Q94LD6	Q94LD6 oryza sativ
30	75	17.9	1081	4 Q9HBF3	Q9HBF3 oryza sativ
31	75	17.9	1992	3 Q9PFT1	Q9PFT1 neopospora
32	74.5	17.8	598	16 Q8YKN7	Q8YKN7 mycobacteri
33	74.5	17.8	894	10 Q9M6S0	Q9M6S0 arabidopsis
34	74.5	17.8	894	10 Q9SUS3	Q9SUS3 arabidopsis
35	74	17.7	384	11 Q9D789	Q9D789 mus musculu
36	73.5	17.6	574	4 Q9BNQ8	Q9BNQ8 homo sapien
37	73.5	17.6	2321	12 Q9DGT6	Q9DGT6 turkey herp
38	73	17.5	123	12 Q9WLK2	Q9WLK2 hepatitis e
39	73	17.5	639	4 Q9H9M1	Q9H9M1 homo sapien
40	73	17.5	1180	5 Q9VWM2	Q9VWM2 drosophila
41	73	17.5	1664	4 Q9BZES	Q9BZES homo sapien
42	73	17.5	3503	5 Q24292	Q24292 drosophila
43	72.5	17.3	229	10 Q94H11	Q94H11 oryza sativ
44	72.5	17.3	518	16 Q3738	Q3738 mycobacteri
45	72.5	17.3	715	2 Q9F9V7	Q9F9V7 mycobacteri

## ALIGNMENTS

RESULT 1	ID	Q9UK79	PRELIMINARY:	PRT:	419 AA.
AC	Q9UK79:				
DT	01-MAY-2000 (TREMBLrel. 13, Created)				
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Herstatin.				
GN	HER-2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-99415951; PubMed-10485918;				
RA	Doherty J.K., Bond C., Jaridim A., Adelman J.P., Clinton G.M.;				
RT	"The HER-2/neu receptor tyrosine kinase gene encodes a secreted				
RT	autoinhibitor.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;				
RL	Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.				
DR	EMBL; AF177761; AAD56009.2; -				
DR	InterPro: IPR000494; EGFR_L-domain.				
DR	InterPro: IPR002174; Furin-like.				
DR	Pfam: PF00757; Furin-like; 1.				
DR	Pfam: PF01030; Recep_L-domain; 1.				
DR	SMART; SM00261; FU; 1.				
SQ	SEQUENCE 419 AA: 45472 MW; FECLBE347E2D030C CRC64;				
Query Match	97.1%; Score 406; DB 4; Length 419;				
Best local Similarity	97.5%; Pred. No. 2.2e-35;				
Matches	77; Conservative 1; Mismatches 1; Indels 0; Gaps 0;				
OY	1 GTHSLPRPAAVVPLRMQGPRAHNVLSFLRPSWDLVSAFYSLPLAPLPTSVTPSPVSV 60				
DB	341 GTHSLPRPAAVVPLRMQGPRAHNVLSFLRPSWDLVSAFYSLPLAPLPTSVTPSPVSV 400				

```
QY      61 GRGPPDAHVAVNLSRYEG 79
          |||||
Db      401 GRGPPDAHVAVDLSRYEG 419
```

## RESULT 2

ID	08UZB4	PRELIMINARY	PRT;	309 AA.
AC	08UZB4			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, last sequence update)			
DT	01-MAR-2002 (TrEMBLrel. 20, last annotation update)			
DE	Hypothetical 31.5 kDa protein.			
OS	Grapevine fleck virus.			
OC	Viruses; unclassified viruses.			
OX	NCBI_TaxID=103722;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT48;			
RA	Sabanadzovic S;			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases			

Query Match	23.8%	Score	99.5%	DB	12	Length	309
Best Local Similarity	39.0%	Pred. NO	0.0076				
Matches	30	Conservative	8	Mismatches	20	Indels	19
						Gaps	5

```

QY      6  LKPKAAVV-----PLKMQGPG-ANGVLS-----FLKPSMDVLSAAYSLIAPLSPTS 52
Db      176 VPRSAVAPFLPLARELPPRLRPLPAPRYLHPLRLARLRLRPPEDLSP---PLSP--PLS 225
QY      53 VPISPSVSGRGPPDPAH 69
Db      230 PRLSPISPLNAPRPPH 245

```

### RESULT 3

ID	Q99JK6.	PRELIMINARY;	PRT:	327 AA.
AC	Q99JK6.			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DY	01-JUN-2001 (TREMBLrel. 17, last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)			
DE	Hypothetical 33.7 kDa protein (Fragment).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae			
OX	NCBI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Straussberg R.;			
RL	Submitted (Apr-2001) to the EMBL/GenBank/DDay databases.			
DR	EMBL; BC006054; AAB06054.1; -			
DR	InterPro: IPR002965; P_Rich_extensn.			
DR	PRINTS: PR01217; PRICHEXTENSX.			
KW	Hypothetical protein.			
FT	NON_TER	1		
SQ	SEQUENCE	327 AA;	33661 MW;	27917FI6D583E774 CRC64;
Query Match		20.3%;	Score 85;	DB 11; Length 327;
Best Local Similarity		42.4%;	Pred. No. 0.28;	

Matches	28; Conservative	4; Mismatches	28; Indels	6; Gaps	4
QY	1	GHHLLPRANPVPLRMQPGRAHVLSPLRSMWDLSAFYSLPLAPLSPTSPVSPVS	60		
Db	88	GAGSGSSP-PGRCPALRQDSTPTNPAWSPRRPAKGLDAA--SSP--PLEPGSPRPSF-PA	141		
QY	61	GGGPPD	66		
Db	142	GLSPSP	147		

## RESULT 4

ID	09GCU15	PRELIMINARY;	PRT;	434 AA.
AC	09GCU15			
DT	01-MAR-2001	(TREMBLrel_16, Created)		
DT	01-MAR-2001	(TREMBLrel_16, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel_19, Last annotation update)		
DE	Hypothetical 49.7 kDa protein.			
OS	Theileria parva.			
OC	Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileridae;			
OC	Theileria.			
OX	NCBI_TaxID=5875;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	Biochem	20520968	PubMed=11071289;	
RA	Bishop R., Gohright E., Nene V., Morzaria S., Musoke A., Schapach B., et al.			
RT	"Polymorphic open reading frames encoding secretory proteins are located less than 3 kilobases from Theileria parva telomeres.";			
RL	Mo. Blochem. Parasitol.	110:359-371(2000).		
DR	EMBL;	AF225701;	AAC80822.1;	
DR	InterPro;	IPR002965;	P-rich_extensn.	
DR	PRINTS;	PR01217;	PRICHEXTENSN.	
KM	Hypothetical protein.			
SQ	SEQUENCE	434 AA;	49670 MW;	0F46586F4021A22F CRC64;

Query Match	20.28;	Score	84.5;	DB	5;	Length	434;
Best Local Similarity	31.0%;	Pred.	No. 0.43;				
Matches	26;	Conservative	11;	Mismatches	34;	Indels	13;
						Gaps	4

QY 4 SLLEPAAVVY-~~PLRM~~PG-~~PAHV~~LSLEKPSMDLSAFA-----YSLPLAPLSPTSVP 55  
 Db 159 ALYPPPTQPTQPTPLPLPGHHPARI-----PRYVQSGYSFYHRYARPPSPVQIPPPS 213  
 QY 56 SPVSYGRGRRPDDPAHVAVNLSRREG 79  
 Db 214 THYVPTPTQPQRPQVPRQVQYTG 237

## RESULT 5

ID	035615:	PRELIMINARY:	PRT:	995 AA.
AC	035615:			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	FOG.			
CN	ZFPML OR FOG.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus			
OX	[NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97373824; PubMed=9230307;			
RA	Tseng A.P., Visvader J.E., Turner C.A., Fujiwara Y., Yu C.,			
RA	Wells M.J., Crossley M., Oskin S.H.;			
RT	"FOG, a multitype zinc finger protein, acts as a cofactor for			
RT	transcription factor GATA-1 in erythroid and megakaryocytic			
RT	differentiation.";			
RL	Cell 90:109-119(1997).			
CC	-i- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).			
DR	EMBL; AF006492; AAC53292.1; -			



Db 112 V 112

RESULT 9

ID 070474 PRELIMINARY; PRT; 816 AA.

AC 070474;

DT 01-AUG-1998 (TReMBLrel. 07, Created)

DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Neurocan (Fragment).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SPRAGUE-DAWLEY; TISSUE=HIPPOCAMPUS;

RA Zachmann-Brand B., Schaller H.C.;

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF060879; AAC15766.1; -.

DR HSSP: P00740; 1EDM.

DR InterPro: IPR000152; Asx\_hydroxyl.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR000742; EGF\_2.

DR InterPro: IPR001881; EGF\_Ca.

DR InterPro: IPR001438; EGF\_II.

DR InterPro: IPR001304; Lectin\_C.

DR InterPro: IPR000538; Link.

DR Pfam: PF00008; EGF\_2.

DR Pfam: PF00059; Lectin\_C; 1.

DR Pfam: PF00193; Xlink; 1.

DR PRINTS: PRO0010; EGFBL00D.

DR PRINTS: PRO1265; LINKMODULE.

DR ProDom: PD000918; Link; 1.

DR SMART: SM00179; EGF\_CA; 1.

DR SMART: SM00001; EGF\_Like; 1.

DR SMART: SM00445; Link; 1.

DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.

DR PROSITE: PS00041; C\_TYPE\_LLECTIN\_2; 1.

DR PROSITE: PS00022; EGF\_1; UNKNOWN\_3.

DR PROSITE: PS01186; EGF\_2; 1.

DR PROSITE: PS01187; EGF\_CA; 1.

DR PROSITE: PS01241; Link; 1.

KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.

FT NON\_TER 1

FT NON\_TER 816

SQ SEQUENCE 816 AA; 85578 MW; 957F5917AD10616E CRC64;

Query Match 19.5%; Score 81.5; DB 11; Length 816;

Best local Similarity 35.0%; Pred. No. 1.7;

Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

QY 4 SLLEPPAAVPLRMQPG---PAHPVLSFLR-----PSMDLVSAFYSLPLAPLS--PT 51

Db 342 SSIPSESLASVLSQASPDGSDPFPVAMLRAPKTLWLPSTLVPPNVSPILPSAPLPS 401

QY 52 SVP-----ISPVSYGRGDPD 67

Db 402 SYPEQAVRVSFG-AEDPE 420

RESULT 10

ID 082761 PRELIMINARY; PRT; 200 AA.

AC 082761;

DT 01-NOV-1998 (TReMBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Putative phytoeyanin (Putative blue copper-binding protein).

GN ATG31050.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RX MEDLINE-20083487; PubMed-10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V., Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S., Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R., RACopenhaver G.P., Preuss D., Nierman W.C., White O., Eissen J.A., Salberg S.L., Fraser C.M., Venter J.C.;

RA \*Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.\*

RT thaliana.

RL Nature 402:761-768(1999).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Rounsley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R., Venter J.C.;

RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

RL [4]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Town C.D., Kaul S.;

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC005311; AAC63847.1; -.

DR EMBL: AC004669; AAM14981.1; -.

DR HSSP: P00303; 2CBP.

DR InterPro: IPR003245; Plectanin\_Like.

DR Pfam: PF02298; Cu\_bind\_Like; 1.

DR ProDom: PD003122; Plectanin\_Like; 1.

SQ SEQUENCE 200 AA; 21475 MW; E669011C997E349C CRC64;

Query Match 19.4%; Score 81; DB 10; Length 200;

Best local Similarity 34.3%; Pred. No. 0.45;

Matches 23; Conservative 5; Mismatches 23; Indels 16; Gaps 2;

QY 9 PAAVPLRMQPGPAHPVLSFLRPSMDLVSAFYSLPLAPISPTSPVSYGRGDPDA 68

Db 132 PVAAPV-----PGVVRPSSFSFPS-----QSPLAESPVNHAPVQYQMGSPAP 175

QY 69 HVAVNLS 75

Db 176 HSAASNS 182

RESULT 11

ID 091M01 PRELIMINARY; PRT; 1006 AA.

AC 091M01;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE F7H2.17 protein.

GN F7H2.17.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Liu S.X., Sakano H., Yu G., Etgu P., Lee J., Lenz C., Pham P.,

[illegible]

```

RN [5] SEQUENCE FROM N.A.
RP STRAIN=LCL8664;
RX MEDLINE=21602573; PubMed=11739708;
RA Rivaller P., Jang H., Cho Y.-G., Quink C., Wang F.;
RT "Complete Nucleotide Sequence of the Rhesus Lymphocryptovirus: Genetic
RL J. Virol. 76:421-426(2002).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Moghaddam A., Koch J., Anis B., Wang F.;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Moghaddam A., Anis B., Wang F.;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Rivaller P., Quink C., Wang F.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Jang H., Wang F.;
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Rao P.V., Jang H., Wang F.;
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=LCL8664;
RA Rivaller P., Jang H., Cho Y.-G., Quink C., Wang F.;
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY037858; AAF7881.2; -
SQ SEQUENCE 941 AA; 103089 MW; EA4E3DC9BEC19A92 CRC64;

Query Match 18.8%; Score 78.5; DB 12; Length 941;
Best Local Similarity 31.9%; Pred. No. 4.2;
Matches 22; Conservative 8; Mismatches 16; Indels 23; Gaps

QY 7 PRPAAPVPLRMQCPAPAPV-----TSFLRPSMDIVSAFSLPLAPLSPSY-PISPY 58
Db 580 PRPAAPL-SLALIGPAPETSSSVSYAAPS-----APIPQAPVPI 624
QY 59 SVGRGPPD 67
Db 625 PIRCGPPD 633

RESULT 13
Q9YLRO PRELIMINARY; PRT; 122 AA.
AC Q9YLRO;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical 12.4 kDa protein.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US2;
RA Schlauder G.G., Dawson G.J., Erker J.C., Kwo P.Y., Knigge M.F.,
RA Shalley D.L., Rosenblatt J.E., Desai S.M., Mushahwar I.K.;
RT "The sequence and phylogenetic analysis of a novel hepatitis E virus

```

```
QY      60 VGRGPPDA 68
          | | |
Db      69 ----PAPPA 73
```

Db	121	PAPAIAPISTIAQVPVMPQPSVTPFAVSPAAAPANAVAPAPAPAPAPAPAPVVPVAPAPAS	160
Qy	52	SVPISPVS	59

QY	52	SVPISPV	59
		::  :	
Db	181	VVPVAPVA	188

Tue Mar 4 13:30:20 2003

us-09-234-208b-1.rspt

Page 7

Search completed: March 4, 2003, 12:35:58  
job time : 10.7249 secs

---





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:32:36 ; Search time 4.28313 Seconds  
(without alignments)  
542.689 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 418  
Sequence: 1 GTHSLPRPAVPVPLRMQP.....VGRGPPDAHVAVNLSTREG 79

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PT05\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	418	100.0	79	4	US-09-630-155-1
2	418	100.0	419	4	US-09-630-155-2
3	81.5	19.5	1257	1	US-08-340-428B-49
4	77	18.4	122	4	US-09-462-606-65
5	72	17.2	440	4	US-08-430-286A-9
6	71.5	17.1	604	2	US-08-468-576B-12
7	71.5	17.1	604	2	US-08-468-576B-12
8	71.5	17.1	604	3	US-08-468-576B-12
9	71.5	17.1	2441	1	US-08-194-468-2
10	71.5	17.1	2441	3	US-08-961-739-2
11	71.5	17.1	2441	4	US-09-514-247A-8
12	69.5	16.6	344	4	US-09-147-236-11
13	69.5	16.6	432	1	US-08-615-170-21
14	69.5	16.6	432	1	US-08-615-170-21
15	68.5	16.4	415	4	US-09-615-170-19
16	68.5	16.4	132	4	US-09-461-697-58
17	68.5	16.4	139	4	US-09-461-697-54
18	68.5	16.4	159	4	US-09-461-697-52
19	68	16.3	2972	4	US-09-579-181-2
20	68	16.3	3118	4	US-09-579-181-1
21	67.5	16.1	123	3	US-08-840-316-3
22	67.5	16.1	123	3	US-08-478-507-9
23	67.5	16.1	123	4	US-08-809-523-3
24	67.5	16.1	123	4	US-09-128-275A-9
25	67.5	16.1	123	4	US-08-471-971-3
26	67.5	16.1	123	4	US-09-553-427-9
27	67.5	16.1	123	4	US-09-462-606-13

28	67.5	16.1	123	4	US-09-462-606-59	Sequence 59, Appl
29	67.5	16.1	123	4	US-09-462-606-60	Sequence 60, Appl
30	67.5	16.1	123	4	US-09-462-606-62	Sequence 62, Appl
31	67.5	16.1	123	4	US-09-462-606-63	Sequence 63, Appl
32	67.5	16.1	123	4	US-09-462-606-64	Sequence 64, Appl
33	67.5	16.1	123	4	US-09-402-776-3	Sequence 3, Appl1
34	67.5	16.1	123	5	PCR-US93-08849A-3	Sequence 3, Appl1
35	67.5	16.1	123	5	PCR-US93-08849-3	Sequence 3, Appl1
36	67.5	16.1	124	1	US-08-240-049B-19	Sequence 19, Appl
37	67.5	16.1	124	4	US-08-542-634-21	Sequence 21, Appl
38	67.5	16.1	124	4	US-08-477-292-21	Sequence 21, Appl
39	67.5	16.1	124	5	PCR-US95-13703-21	Sequence 21, Appl
40	67.5	16.1	136	4	US-08-259-451-5	Sequence 5, Appl1
41	67.5	16.1	433	4	US-08-259-451-3	Sequence 3, Appl1
42	67	16.0	174	4	US-08-818-112-143	Sequence 143, App
43	67	16.0	174	4	US-08-818-111-138	Sequence 138, App
44	67	16.0	174	4	US-09-056-556-143	Sequence 143, App
45	67	16.0	174	4	US-09-072-596-138	Sequence 138, App

## ALIGNMENTS

```
RESULT 1
US-09-630-155-1
: Sequence 1, Application US/09630155
: Patent No. 6A14130
:
: GENERAL INFORMATION:
: APPLICANT: Doherty, Joni Kristin and Gail M. Clinton
: TITLE OF INVENTION: HER-2 BINDING ANTAGONISTS
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DAVIS WRIGHT TREMAINE LLP
: STREET: 1501 Fourth Avenue, 2600 Century Square
: CITY: Seattle
: STATE: Washington
: COUNTRY: U.S.A.
: ZIP: 98101
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: PC compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: Word
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/630,155
: FILING DATE: 16-Jan-2001
: CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Davison, Barry L.
: REGISTRATION NUMBER: 47,309
: REFERENCE/DOCKET NUMBER: 49321-10
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206 628-7621
: TELEFAX: 206 628-7699
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 79
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: HER-2 ECD antagonist
: SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-630-155-1

Query Match          100.0%; Score 418; DB 4; Length 79;
Best Local Similarity 100.0%; Pred. No. 5.2e-42;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GTHSLPRPAVPVPLRMQPGPAHVPVLSFLRPSMDVSAFTSLPLAPISPTSPVPSV 60
Db 1 GTHSLPRPAVPVPLRMQPGPAHVPVLSFLRPSMDVSAFTSLPLAPISPTSPVPSV 60
Oy 61 GRGPPDAHVAVNLSTREG 79
```

Db 61 GRGPPDAHVAVNLRYEG 79

## RESULT 2

US-09-630-155-2  
Sequence 2, Application US/09630155  
Patent No. 6414130

## GENERAL INFORMATION:

APPLICANT: Doherty, Joni Kristin and Gail M. Clinton  
TITLE OF INVENTION: HER-2 BINDING ANTAGONISTS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVIS WRIGHT TREMAINE LLP  
STREET: 1501 Fourth Avenue, 2600 Century Square  
City: Seattle  
STATE: Washington  
COUNTRY: U.S.A.

ZIP: 98101

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: PC compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: Word

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/630,155  
FILING DATE: 16-Jan-2001  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Davison, Barry L.  
REGISTRATION NUMBER: 47,309  
REFERENCE/DOCKET NUMBER: 49321-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206 628-7621  
TELEFAX: 206 628-7693

## INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 419  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown

MOLECULE TYPE: polypeptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-630-155-2

Query Match 100.0%: Score 418; DB 4; Length 419;  
Best Local Similarity 100.0%: Pred. NO. 4.2e-41;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTHSLPRPAVPVPLRMQPGPAHPLSLFRSPWDLVSFAFVSLPLAPLSPTSPVSPVSV 60

Db 341 GTHSLPRPAVPVPLRMQPGPAHPLSLFRSPWDLVSFAFVSLPLAPLSPTSPVSPVSV 400

OY 61 GRGPPDAHVAVNLRYEG 79

Db 401 GRGPPDAHVAVNLRYEG 419

## RESULT 3

US-08-340-428B-49  
Sequence 49, Application US/08340428B  
Patent No. 5648465

## GENERAL INFORMATION:

APPLICANT: MARGOLIS, Richard U.  
APPLICANT: RAUCH, Uwe

APPLICANT: MARGOLIS, Renee K.

TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A  
TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W.  
CITY: Washington

STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/340,428B  
FILING DATE: 14 No. 5648465ember 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/922,911  
FILING DATE: 03 August 1992  
CLASSIFICATION: 514

## ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: Margolis=1A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528

## INFORMATION FOR SEQ ID NO: 49:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-340-428B-49

Query Match 19.5%: Score 81.5; DB 1; Length 1257;  
Best Local Similarity 35.0%: Pred. NO. 0.38;  
Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

OY 4 SLPRPAVPVPLRMQPG---PAHPVLSFLR-----PSMDLSAFSLPLAPLS--PT 51

Db 610 SLPRPAVPVPLRMQPG---PAHPVLSFLR-----PSMDLSAFSLPLAPLS--PT 669

OY 52 SVP---ISPVSGRPPD 67

Db 670 SVP---ISPVSGRPPD 688

## RESULT 4

US-09-462-606-65  
Sequence 65, Application US/09462606  
Patent No. 6432408

## GENERAL INFORMATION:

APPLICANT: MENG, XIANG-JIN  
APPLICANT: Emerson, Suzanne U.

APPLICANT: Purcell, Robert H.

TITLE OF INVENTION: A SWINE HEPATITIS E VIRUS AND USES THEREOF  
FILE REFERENCE: 20264267051

CURRENT APPLICATION NUMBER: US/09/462,606

CURRENT FILING DATE: 2000-06-12

PRIOR APPLICATION NUMBER: US 60/053069

PRIOR FILING DATE: 1997-07-18

PRIOR APPLICATION NUMBER: PCT/US98/14665

PRIOR FILING DATE: 1998-07-17

NUMBER OF SEQ ID NOS: 65

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 65

LENGTH: 122

TYPE: PRT

ORGANISM: Hepatitis E virus

US-09-462-606-65

Query Match 18.4%: Score 77; DB 4; Length 122;  
Best Local Similarity 36.1%: Pred. NO. 0.067;  
Matches 22; Conservative 10; Mismatches 23; Indels 6; Gaps 3;



COUNTRY: USA  
ZIP: 10591-5144  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: System 7.5  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,579B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,276  
FILING DATE: 05-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/872,646  
FILING DATE: 08-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/715,181  
FILING DATE: 14-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/441,703  
FILING DATE: 04-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/312,543  
FILING DATE: 17-FEB-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurt G. Briscoe  
REGISTRATION NUMBER: 33,141  
REFERENCE/DOCKET NUMBER: MDI 251.5-KGB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 332-1700  
TELEFAX: (914) 332-1844  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 604 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-468-579B-12

Query Match 17.1%; Score 71.5; DB 2; Length 604;  
Best Local Similarity 34.8%; Pred. No. 2.2;  
Matches 23; Conservative 13; Mismatches 27; Indels 3; Gaps 2;

QY 10 AAVPVPLRMQPGPAHVLSFLRPSMDLVSAFVSLPLAPLSPVSPISPV--SVGRGPPDP 67  
DB 421 AAOPLMSSRPKTAEPVKPTPTONLFPASKTSPVNLPMKSSIP-SPIGSLGRSSLD 479

QY 68 AHVAVN 73  
DB 480 ILSSLN 485

RESULT 8  
US-08-468-577B-12  
Sequence 12, Application US/08468577B  
Patent No. 6001804  
GENERAL INFORMATION:  
APPLICANT: Rabin, Daniel  
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS  
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sprung Kramer Schaefer & Briscoe  
STREET: 660 White Plains Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591-5144  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: System 7.5

SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,577B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,276  
FILING DATE: 05-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/872,646  
FILING DATE: 08-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/715,181  
FILING DATE: 14-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/441,703  
FILING DATE: 04-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/312,543  
FILING DATE: 17-FEB-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurt G. Briscoe  
REGISTRATION NUMBER: 33,141  
REFERENCE/DOCKET NUMBER: MDI 251.8-KGB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 332-1700  
TELEFAX: (914) 332-1844  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 604 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-468-577B-12

Query Match 17.1%; Score 71.5; DB 3; Length 604;  
Best Local Similarity 34.8%; Pred. No. 2.2;  
Matches 23; Conservative 13; Mismatches 27; Indels 3; Gaps 2;

QY 10 AAVPVPLRMQPGPAHVLSFLRPSMDLVSAFVSLPLAPLSPVSPISPV--SVGRGPPDP 67  
DB 421 AAOPLMSSRPKTAEPVKPTPTONLFPASKTSPVNLPMKSSIP-SPIGSLGRSSLD 479

QY 68 AHVAVN 73  
DB 480 ILSSLN 485

RESULT 9  
US-08-194-468-2  
Sequence 2, Application US/08194468  
Patent No. 5750336  
GENERAL INFORMATION:  
APPLICANT: Montminy, Marc R.  
TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF  
TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/194,468  
FILING DATE: 10-FEB-1994  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9672  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619)-546-4737  
TELEFAX: (619)-546-9392  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2441 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-194-468-2

Query Match 17.1%; Score 71.5; DB 1; Length 2441;  
Best Local Similarity 36.5%; Pred. No. 13;

Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

QY 4 SLPRPAVPVPLRMQPGAHVPLSLRPSWDLVSIFYSLPLAPLSPTSVIPSVGK 63

DB 843 SQLPCPPVQSLHPTPPASTAGM--PSLQHTAPGHTPPQAPATQ-PSIPVSSGQT 899

QY 64 PDP 66

DB 900 PTP 902

## RESULT 10

US-08-961-739-2  
Sequence 2, Application US/08961739A  
Patent No. 6063583

GENERAL INFORMATION:

APPLICANT: Montminy, Marc R.  
TITLE OF INVENTION: Methods for Treating Diabetes Mellitus

FILE REFERENCE: SALK1650-1

CURRENT APPLICATION NUMBER: US/08/961,739A

CURRENT FILING DATE: 1997-10-31

EARLIER APPLICATION NUMBER: US 194,468

EARLIER FILING DATE: 1994-02-10

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 2441

TYPE: PRT

ORGANISM: Mus

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)...(2441)

OTHER INFORMATION: Xaa = Any Amino Acid

US-08-961-739-2

Query Match 17.1%; Score 71.5; DB 3; Length 2441;  
Best Local Similarity 36.5%; Pred. No. 13;

Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

QY 4 SLPRPAVPVPLRMQPGAHVPLSLRPSWDLVSIFYSLPLAPLSPTSVIPSVGK 63

DB 843 SQLPCPPVQSLHPTPPASTAGM--PSLQHTAPGHTPPQAPATQ-PSIPVSSGQT 899

QY 64 PDP 66

DB 900 PTP 902

## RESULT 11

US-09-514-247A-8  
Sequence 8, Application US/09514247A  
Patent No. 6365361

GENERAL INFORMATION:

APPLICANT: TANABE SEIYAKU CO. LTD.

APPLICANT: TANIGUCHI, Tomoyasu

APPLICANT: MIZUKAMI, Junko

TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO  
FILE REFERENCE: TANIGUCHI-6  
CURRENT APPLICATION NUMBER: US/09/514,247A  
CURRENT FILING DATE: 2000-02-28  
PRIOR APPLICATION NUMBER: PCT/JP98/03734  
PRIOR FILING DATE: 1998-08-24  
PRIOR APPLICATION NUMBER: JP231084/1997  
PRIOR FILING DATE: 1997-08-27  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 8  
LENGTH: 2441  
TYPE: PRT  
ORGANISM: mouse  
US-09-514-247A-8

Query Match 17.1%; Score 71.5; DB 4; Length 2441;  
Best Local Similarity 36.5%; Pred. No. 13;

Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

QY 4 SLPRPAVPVPLRMQPGAHVPLSLRPSWDLVSIFYSLPLAPLSPTSVIPSVGK 63

DB 843 SQLPCPPVQSLHPTPPASTAGM--PSLQHTAPGHTPPQAPATQ-PSIPVSSGQT 899

QY 64 PDP 66

DB 900 PTP 902

## RESULT 12

US-09-147-236-11  
Sequence 11, Application US/09147236A  
Patent No. 6316251

GENERAL INFORMATION:

APPLICANT: TONOUCHI, Naoto

APPLICANT: TSUCHIDA, Takayasu

APPLICANT: YOSHINAGA, Fumihito

APPLICANT: TAHARA, Naoki

APPLICANT: HAYASHI, Takahisa

TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE

FILE REFERENCE: 6537-011-0PCT

CURRENT APPLICATION NUMBER: US/09/147,236A

CURRENT FILING DATE: 1999-04-08

EARLIER APPLICATION NUMBER: PCT/JP97/03633

EARLIER FILING DATE: 1997-10-09

NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 11

LENGTH: 344

TYPE: PRT

ORGANISM: Acetobacter xylinum

FEATURE:

OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, c, or

OTHER INFORMATION: t

US-09-147-236-11

Query Match 16.6%; Score 69.5; DB 4; Length 344;  
Best Local Similarity 31.0%; Pred. No. 1.9;

Matches 22; Conservative 7; Mismatches 19; Indels 23; Gaps 3;

QY 3 HSLPRPAVPVPLRMQPGAHVPLSLRPSWDLVSIFYSLPLAPLSPTSVIPSVPI 55

DB 158 YAAAPQPVATPPV--PQAPAVAAVAAQPVQKQERS-----LSPTPK 201

QY 56 SPVSVGRGPP 66

DB 202 PAVSFMADRP 112

RESULT 13  
US-08-615-170-21  
Sequence 21, Application US/08615170  
Patent No. 5776776

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/615,170
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01526
FILING DATE: 06-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/191,493
FILING DATE: 04-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 2307U-053120
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-615-170-19

Query Match          16.6%; Score 69.5; DB 1; Length 433;
Best Local Similarity 35.0%; Pred. No. 2.5;
Matches 21; Conservative 11; Mismatches 17; Indels 11; Gaps 4;

OY      4 SLDPPAAVYPVLPMQGPAPHPVLSTLRPSMDLVS-----AFYSLPPLADLSP--TSVP I 55
         | | | : | : | | | | : | : | : | | | | | | | | | | | | | | | | | | |
Db       156 SAARFWSGPIP--GGGPGSQDIKPFQAQPAVYPIQPPMPSPSLASVE--PLAVLPMAASAVPY 212

RESULT 15
US-09-461-697-58
: Sequence 58, Application US/09461697
: Patent No. 6277974
: GENERAL INFORMATION:
: APPLICANT: COGENET NEUROSCIENCE, Inc.
: APPLICANT: LO, Donald C.
: APPLICANT: Barney, Shawn
: APPLICANT: Thomas, Mary Beth
: APPLICANT: Portbury, Stuart D.
: APPLICANT: Puranam, Kasturi
: APPLICANT: Katz, Lawrence C.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
: TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
: TITLE OF INVENTION: CELL DEATH
FILE REFERENCE: 10001-005-999
: CURRENT APPLICATION NUMBER: US/09/461,697
: NUMBER OF SEQ ID NOS: 466
: SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 58
LENGTH: 115
TYPE: PRT
ORGANISM: Homo sapiens

```

US-09-461-697-58

Query Match	16.48;	Score 68.5;	DB 4;	Length 115;
Best Local Similarity	27.8%;	Pred No 0.61;		

Matches 25; Conservative 9; Mismatches 29; Indels 27; Gaps 4;

[illegible]

Search completed: March 4, 2003, 12:37:08  
Job time : 6.28313 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:36:06 ; Search time 3.33133 Seconds  
(without alignments)  
1000.035 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 418  
Sequence: 1 GTHSLIPRAAVPVLRLMQP.....VGRGPDPAHVAVNLRYREG 79

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.5	20.0	888	US-10-036-041-35	Sequence 35, Appl
2	83.5	20.0	888	US-10-028-072-544	Sequence 544, App
3	83.5	20.0	888	US-10-035-855-35	Sequence 35, Appl
4	83.5	20.0	888	US-10-121-049-544	Sequence 544, App
5	83.5	20.0	888	US-10-123-904-544	Sequence 544, App
6	83.5	20.0	888	US-10-140-470-544	Sequence 544, App
7	83.5	20.0	888	US-09-931-836-35	Sequence 35, Appl
8	83.5	20.0	888	US-10-175-746-544	Sequence 544, App
9	83.5	20.0	888	US-10-176-918-544	Sequence 544, App
10	83.5	20.0	888	US-10-176-921-544	Sequence 544, App
11	83.5	20.0	888	US-10-036-214-35	Sequence 35, Appl
12	83.5	20.0	888	US-10-137-865-544	Sequence 544, App
13	83.5	20.0	888	US-10-140-474-544	Sequence 544, App
14	83.5	20.0	888	US-10-035-719-35	Sequence 35, Appl
15	83.5	20.0	888	US-10-142-431-544	Sequence 544, App
16	83.5	20.0	888	US-10-143-114-544	Sequence 544, App
17	83.5	20.0	888	US-10-140-002-544	Sequence 544, App
18	83.5	20.0	888	US-10-036-342-35	Sequence 35, Appl
19	74	17.7	616	US-09-925-300-1519	Sequence 1519, Ap

20	73	17.5	3503	9	US-10-108-605-237	Sequence 237, App
21	72	17.2	440	9	US-09-966-782A-7	Sequence 7, Appl
22	71.5	17.1	2441	12	US-10-109-886-8	Sequence 8, Appl
23	71	17.0	463	9	US-10-029-180-80	Sequence 80, Appl
24	69.5	16.6	802	10	US-09-823-240-2	Sequence 2, Appl
25	69	16.5	265	10	US-09-819-254-2	Sequence 2, Appl
26	69	16.5	265	10	US-09-998-598-2592	Sequence 2592, Ap
27	68.5	16.4	115	10	US-09-922-261-58	Sequence 58, Appl
28	68.5	16.4	132	10	US-09-922-261-54	Sequence 54, Appl
29	68.5	16.4	139	10	US-09-922-261-52	Sequence 52, Appl
30	68.5	16.4	159	10	US-09-922-261-48	Sequence 48, Appl
31	68	16.3	171	10	US-09-925-297-659	Sequence 659, App
32	68	16.3	3298	9	US-10-160-758-16	Sequence 16, Appl
33	67.5	16.1	124	10	US-09-769-066-21	Sequence 21, Appl
34	67.5	16.1	279	9	US-10-001-857-159	Sequence 159, App
35	67.5	16.1	609	10	US-09-796-338A-5	Sequence 5, Appl
36	67.5	16.1	609	10	US-09-782-980-74	Sequence 74, Appl
37	66.5	15.9	105	9	US-09-764-868-1235	Sequence 1235, Ap
38	66.5	15.9	435	10	US-09-749-728B-29	Sequence 29, Appl
39	66	15.8	119	10	US-09-864-761-36209	Sequence 36209, A
40	66	15.8	549	9	US-09-989-442-100	Sequence 100, App
41	66	15.8	744	9	US-09-764-868-667	Sequence 667, App
42	65.5	15.7	218	9	US-10-001-873-35	Sequence 35, Appl
43	65.5	15.7	1114	10	US-09-740-046-10	Sequence 10, Appl
44	64.5	15.4	124	10	US-09-769-066-22	Sequence 22, Appl
45	64.5	15.4	405	10	US-09-803-126-13	Sequence 13, Appl

#### ALIGNMENTS

RESULT 1  
US-10-036-041-35  
Sequence 35, Application US/10036041  
Publication No. US20020192751A1  
GENERAL INFORMATION:  
APPLICANT: Desnoyers, Luc  
APPLICANT: Baton, Dan L.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3030R1C8  
CURRENT FILING DATE: 2001-12-26  
PRIOR APPLICATION NUMBER: US/10/036,041  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/112514  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113300  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/113430  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113605  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113621  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/114140  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/115552  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/116843  
PRIOR FILING DATE: 1999-01-22  
PRIOR APPLICATION NUMBER: 60/125774  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 60/125778  
PRIOR FILING DATE: 1999-03-23

PRIOR APPLICATION NUMBER: 60/125826  
PRIOR FILING DATE: 1999-03-24  
PRIOR APPLICATION NUMBER: 60/127035  
PRIOR FILING DATE: 1999-03-31  
PRIOR APPLICATION NUMBER: 60/127706  
PRIOR FILING DATE: 1999-04-05  
PRIOR APPLICATION NUMBER: 60/129122  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: 60/130359  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: 60/131270  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131272  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131291  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/132371  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/132379  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/132383  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/135750  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: 60/138166  
PRIOR FILING DATE: 1999-06-08  
PRIOR APPLICATION NUMBER: 60/144791  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/146970  
PRIOR FILING DATE: 1999-08-03  
PRIOR APPLICATION NUMBER: 60/162506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 09/311832  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 09/380142  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/644848  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 09/747259  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 09/816744  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 09/854208  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: 09/854280  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: 09/874503  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: 09/869599  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: 09/908, 827  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: PCT/US99/10733  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30720  
PRIOR FILING DATE: 1999-12-22  
PRIOR APPLICATION NUMBER: PCT/US00/05601  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: PCT/US00/05841  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: PCT/US00/14042  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: PCT/US00/15264  
PRIOR FILING DATE: 2000-06-02  
PRIOR APPLICATION NUMBER: PCT/US00/23522  
PRIOR FILING DATE: 2000-08-23  
PRIOR APPLICATION NUMBER: PCT/US00/23328  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: PCT/US00/32678  
PRIOR FILING DATE: 2000-12-01  
PRIOR APPLICATION NUMBER: PCT/US00/34956

PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: PCT/US01/06520  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/17800  
PRIOR FILING DATE: 2001-06-01  
PRIOR APPLICATION NUMBER: PCT/US01/19692  
PRIOR FILING DATE: 2001-06-20  
PRIOR APPLICATION NUMBER: PCT/US01/21066  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: PCT/US01/21735  
PRIOR FILING DATE: 2001-07-09  
NUMBER OF SEQ ID NOS: 80  
SEQ ID NO 35  
LENGTH: 888  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-036-041-35  
Query Match 20.0%; Score 83.5; DA 9; Length 888;  
Best Local Similarity 35.4%; Pred. No. 0.91;  
Matches 28; Conservative 7; Mismatches 33; Indels 11; Gaps 5;  
QY 1 GTH---SLPRPAAVPLRMQPGPAHPVLSFLRP-SWD---LVSAPYSPLAPLSFT 51  
Db 698 GPHDLSGLPPEOTPLPQKRLPTP-HPHPALGPRAMDGHPLLPASASSSLLLAPA 756  
QY 52 SVPISPVSGRGPPDPAHV 70  
Db 757 RAPEQPPAGE-PTPDGRL 774  
RESULT 2  
US-10-028-072-544  
Sequence 544, Application US/10028072  
Publication No. US20030004311A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: Deforge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang  
TITLE OF INVENTION:  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/028, 072  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/049911  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059115  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059117  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059122  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059184  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/0593535	
PRIOR FILING DATE: 1997-09-19	
PRIOR APPLICATION NUMBER: 60/0595888	
PRIOR FILING DATE: 1997-09-19	
PRIOR APPLICATION NUMBER: 60/0598366	
PRIOR FILING DATE: 1997-09-24	
PRIOR APPLICATION NUMBER: 60/0622500	
PRIOR FILING DATE: 1997-10-17	
PRIOR APPLICATION NUMBER: 60/0622855	
PRIOR FILING DATE: 1997-10-17	
PRIOR APPLICATION NUMBER: 60/0622877	
PRIOR FILING DATE: 1997-10-17	
PRIOR APPLICATION NUMBER: 60/062814	
PRIOR FILING DATE: 1997-10-24	
PRIOR APPLICATION NUMBER: 60/062816	
PRIOR FILING DATE: 1997-10-24	
PRIOR APPLICATION NUMBER: 60/0630455	
PRIOR FILING DATE: 1997-10-24	
PRIOR APPLICATION NUMBER: 60/0630822	
PRIOR FILING DATE: 1997-10-31	
PRIOR APPLICATION NUMBER: 60/0631277	
PRIOR FILING DATE: 1997-10-24	
PRIOR APPLICATION NUMBER: 60/0633277	
PRIOR FILING DATE: 1997-10-27	
PRIOR APPLICATION NUMBER: 60/0633329	
PRIOR FILING DATE: 1997-10-27	
PRIOR APPLICATION NUMBER: 60/0633550	
PRIOR FILING DATE: 1997-10-28	
PRIOR APPLICATION NUMBER: 60/0633561	
PRIOR FILING DATE: 1997-10-28	
PRIOR APPLICATION NUMBER: 60/0633704	
PRIOR FILING DATE: 1997-10-29	
PRIOR APPLICATION NUMBER: 60/0633733	
PRIOR FILING DATE: 1997-10-29	
PRIOR APPLICATION NUMBER: 60/0633755	
PRIOR FILING DATE: 1997-10-17	
PRIOR APPLICATION NUMBER: 60/064248	
PRIOR FILING DATE: 1997-11-03	
PRIOR APPLICATION NUMBER: 60/064809	
PRIOR FILING DATE: 1997-11-07	
PRIOR APPLICATION NUMBER: 60/065186	
PRIOR FILING DATE: 1997-11-12	
PRIOR APPLICATION NUMBER: 60/065846	
PRIOR FILING DATE: 1997-11-17	
PRIOR APPLICATION NUMBER: 60/066364	
PRIOR FILING DATE: 1997-11-21	
PRIOR APPLICATION NUMBER: 60/066453	
PRIOR FILING DATE: 1997-11-24	
PRIOR APPLICATION NUMBER: 60/066511	
PRIOR FILING DATE: 1997-11-24	
PRIOR APPLICATION NUMBER: 60/066770	
PRIOR FILING DATE: 1997-11-24	
PRIOR APPLICATION NUMBER: 60/069212	
PRIOR FILING DATE: 1997-12-11	
PRIOR APPLICATION NUMBER: 60/069278	
PRIOR FILING DATE: 1997-12-11	
PRIOR APPLICATION NUMBER: 60/069334	
PRIOR FILING DATE: 1997-12-11	
PRIOR APPLICATION NUMBER: 60/069694	
PRIOR FILING DATE: 1997-12-16	
PRIOR APPLICATION NUMBER: 60/072320	
PRIOR FILING DATE: 1998-01-23	
PRIOR APPLICATION NUMBER: 60/073612	
PRIOR FILING DATE: 1998-02-04	
PRIOR APPLICATION NUMBER: 60/074086	
PRIOR FILING DATE: 1998-02-09	
PRIOR APPLICATION NUMBER: 60/074092	
PRIOR FILING DATE: 1998-02-09	
PRIOR APPLICATION NUMBER: 60/077791	

PRIOR FILING DATE: 1998-03-12	PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20	PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25	PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-02-27	PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27	PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31	PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09	PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09	PRIOR APPLICATION NUMBER: 60/081695
PRIOR FILING DATE: 1998-04-14	PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24	PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28	PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/085149
PRIOR FILING DATE: 1998-05-12	PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13	PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13	PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13	PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/085697
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/086414
PRIOR FILING DATE: 1998-05-22	PRIOR APPLICATION NUMBER: 60/086430
PRIOR FILING DATE: 1998-05-22	PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28	PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04	PRIOR APPLICATION NUMBER: 60/088730
PRIOR FILING DATE: 1998-06-10	PRIOR APPLICATION NUMBER: 60/088741
PRIOR FILING DATE: 1998-06-10	PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10	PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11	PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17	PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17	PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18	PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19	PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23	

PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090538  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090863  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/091360  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091519  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07

Query Match 20.0%; Score 83.5; DB 9; Length 888;  
Best Local Similarity 35.4%; Pred. No. 0.91;  
Matches 28; Conservative 7; Mismatches 33; Indels 11; Gaps 5;

Oy 1 GTH---SLPRPAVPPVLRMGPAHPVLSFLRP-SMD---LVSIFYSLPLASPT 51  
Db 698 GPHDLSGLPPPEOTPLPKRLPTP-HPHPALGPRAMDHGHPULPASASSLLLLAPA 756

Oy 52 SVTSPVSVGRCPDPDAHV 70  
Db 757 RAPEQPPAPAGE-PTPDGRL 774

RESULT 3  
US-10-035-855-35  
Sequence 35, Application US/10035855  
Publication No. US20030008348A1  
GENERAL INFORMATION:  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3030R1C4  
CURRENT APPLICATION NUMBER: US/10/035,855  
CURRENT FILING DATE: 2001-12-26  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/112514  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113300  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/113430  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113605  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113621  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/114140  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/115552  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/116843  
PRIOR FILING DATE: 1999-01-22  
PRIOR APPLICATION NUMBER: 60/125774  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 60/125778  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 60/125826  
PRIOR FILING DATE: 1999-03-24  
PRIOR APPLICATION NUMBER: 60/127035

PRIOR FILING DATE: 1999-03-31  
PRIOR APPLICATION NUMBER: 60/127706  
PRIOR FILING DATE: 1999-04-05  
PRIOR APPLICATION NUMBER: 60/129122  
PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: 60/130359  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: 60/131270  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131272  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131291  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/132371  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/132379  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/132383  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/135750  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: 60/138166  
PRIOR FILING DATE: 1999-06-08  
PRIOR APPLICATION NUMBER: 60/144791  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/146970  
PRIOR FILING DATE: 1999-08-03  
PRIOR APPLICATION NUMBER: 60/162506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 09/311832  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 09/380142  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/644848  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 09/747259  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 09/816744  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 09/854208  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: 09/854280  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: 09/874503  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: 09/869599  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: 09/908,827  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: PCT/US99/10733  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30720  
PRIOR FILING DATE: 1999-12-22  
PRIOR APPLICATION NUMBER: PCT/US00/05601  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: PCT/US00/05841  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: PCT/US00/14042  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: PCT/US00/15264  
PRIOR FILING DATE: 2000-06-02  
PRIOR APPLICATION NUMBER: PCT/US00/23522  
PRIOR FILING DATE: 2000-08-23  
PRIOR APPLICATION NUMBER: PCT/US00/23328  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: PCT/US00/32678  
PRIOR FILING DATE: 2000-12-01  
PRIOR APPLICATION NUMBER: PCT/US00/34956  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: PCT/US01/06520  
PRIOR FILING DATE: 2001-02-28



```

: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P33030R1C160
: CURRENT APPLICATION NUMBER: US/10/140,470
: CURRENT FILING DATE: 2002-05-06
: Prior Application removed - See Palm or File Wrapper
: NUMBER OF SEQ ID NOS: 550
: SEQ ID NO 544
: LENGTH: 888
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-10-140-470-544

Query Match          20.0%; Score 83.5; DB 9; Length 888;
Best Local Similarity 35.4%; Pred. No. 0.91;
Matches 28; Conservative 7; Mismatches 33; Indels 11; Gaps 5;

OY 1 GTH----SLPRPAVVPVLRMQGPRAHVLSFLRP-SMD---LVSAFYSLAPLSPT 51
DB 698 GPHDLSGLPRTPEQTLPOKRLPTR-HNPHALGPRAMDGHPLLPASASSLLLLAPA 756
OY 52 SVPISPVSGRGPPDANV 70
DB 757 RAPEQPPAPGE-PTPDGRL 774

RESULT 7
US-09-931-836-35
: Sequence 35, Application US/09931836
: Publication No. US20030027249A1
: GENERAL INFORMATION:
: APPLICANT: Desnoyers, Luc
: APPLICANT: Eaton, Dan L.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3030R1C1
: CURRENT APPLICATION NUMBER: US/09/931,836
: CURRENT FILING DATE: 2001-08-16
: PRIOR APPLICATION NUMBER: 60/085579
: PRIOR FILING DATE: 1998-05-15
: PRIOR APPLICATION NUMBER: 60/112514
: PRIOR FILING DATE: 1998-12-15
: PRIOR APPLICATION NUMBER: 60/113300
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: 60/113430
: PRIOR FILING DATE: 1998-12-23
: PRIOR APPLICATION NUMBER: 60/113605
: PRIOR FILING DATE: 1998-12-23
: PRIOR APPLICATION NUMBER: 60/113621
: PRIOR FILING DATE: 1998-12-23
: PRIOR APPLICATION NUMBER: 60/114140
: PRIOR FILING DATE: 1998-12-23
: PRIOR APPLICATION NUMBER: 60/115552
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/116843
: PRIOR FILING DATE: 1999-01-22
: PRIOR APPLICATION NUMBER: 60/125774
: PRIOR FILING DATE: 1999-03-23
: PRIOR APPLICATION NUMBER: 60/125778
: PRIOR FILING DATE: 1999-03-23
: PRIOR APPLICATION NUMBER: 60/125826
: PRIOR FILING DATE: 1999-03-24
: PRIOR APPLICATION NUMBER: 60/127035
: PRIOR FILING DATE: 1999-03-31
: PRIOR APPLICATION NUMBER: 60/127706
```

```

: PRIOR FILING DATE: 1999-04-05
: PRIOR APPLICATION NUMBER: 60/129122
: PRIOR FILING DATE: 1999-04-13
: PRIOR APPLICATION NUMBER: 60/130359
: PRIOR FILING DATE: 1999-04-21
: PRIOR APPLICATION NUMBER: 60/131270
: PRIOR FILING DATE: 1999-04-27
: PRIOR APPLICATION NUMBER: 60/131272
: PRIOR FILING DATE: 1999-04-27
: PRIOR APPLICATION NUMBER: 60/131291
: PRIOR FILING DATE: 1999-04-27
: PRIOR APPLICATION NUMBER: 60/132371
: PRIOR FILING DATE: 1999-05-04
: PRIOR APPLICATION NUMBER: 60/132379
: PRIOR FILING DATE: 1999-05-04
: PRIOR APPLICATION NUMBER: 60/132383
: PRIOR FILING DATE: 1999-05-04
: PRIOR APPLICATION NUMBER: 60/135750
: PRIOR FILING DATE: 1999-05-25
: PRIOR APPLICATION NUMBER: 60/138166
: PRIOR FILING DATE: 1999-06-08
: PRIOR APPLICATION NUMBER: 60/144791
: PRIOR FILING DATE: 1999-07-20
: PRIOR APPLICATION NUMBER: 60/146970
: PRIOR FILING DATE: 1999-08-03
: PRIOR APPLICATION NUMBER: 60/162506
: PRIOR FILING DATE: 1999-10-29
: PRIOR APPLICATION NUMBER: 09/311832
: PRIOR FILING DATE: 1999-05-14
: PRIOR APPLICATION NUMBER: 09/380142
: PRIOR FILING DATE: 1999-08-25
: PRIOR APPLICATION NUMBER: 09/644848
: PRIOR FILING DATE: 2000-08-22
: PRIOR APPLICATION NUMBER: 09/747259
: PRIOR FILING DATE: 2000-12-20
: PRIOR APPLICATION NUMBER: 09/816744
: PRIOR FILING DATE: 2001-03-22
: PRIOR APPLICATION NUMBER: 09/854208
: PRIOR FILING DATE: 2001-05-10
: PRIOR APPLICATION NUMBER: 09/854280
: PRIOR FILING DATE: 2001-05-10
: PRIOR APPLICATION NUMBER: 09/874503
: PRIOR FILING DATE: 2001-06-05
: PRIOR APPLICATION NUMBER: 09/869599
: PRIOR FILING DATE: 2001-06-29
: PRIOR APPLICATION NUMBER: 09/908,827
: PRIOR FILING DATE: 2001-07-18
: PRIOR APPLICATION NUMBER: PCT/US99/10733
: PRIOR FILING DATE: 1999-05-14
: PRIOR APPLICATION NUMBER: PCT/US99/28551
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/30720
: PRIOR FILING DATE: 1999-12-22
: PRIOR APPLICATION NUMBER: PCT/US00/05601
: PRIOR FILING DATE: 2000-03-01
: PRIOR APPLICATION NUMBER: PCT/US00/05841
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: PCT/US00/14042
: PRIOR FILING DATE: 2000-05-22
: PRIOR APPLICATION NUMBER: PCT/US00/15264
: PRIOR FILING DATE: 2000-06-02
: PRIOR APPLICATION NUMBER: PCT/US00/23522
: PRIOR FILING DATE: 2000-08-23
: PRIOR APPLICATION NUMBER: PCT/US00/23328
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: PCT/US00/32678
: PRIOR FILING DATE: 2000-12-01
: PRIOR APPLICATION NUMBER: PCT/US00/34956
: PRIOR FILING DATE: 2000-12-20
: PRIOR APPLICATION NUMBER: PCT/US01/06520
: PRIOR FILING DATE: 2001-08-28
: PRIOR APPLICATION NUMBER: PCT/US01/17800
: PRIOR FILING DATE: 2001-06-01
```



FILE REFERENCE: P330R1C288  
CURRENT APPLICATION NUMBER: US/10/176,921  
CURRENT FILING DATE: 2002-06-20  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 544  
LENGTH: 888  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-176-921-544

Query Match 20.0%; Score 83.5; DB 9; Length 888;  
Best Local Similarity 35.4%; Pred. No. 0.91;  
Matches 28; Conservative 7; Mismatches 33; Indels 11; Gaps 5;

QY I GTH-----SLLRPAVPPRLMOPRAPHVLSFLRP-SWD-----LYSAFYSLPLAPLSPT 51  
Db 698 GPHDIDSGLLPPEQTPYQKRLETPR-HPHPALGPRAMDHGHPLLPASASSLLLLAPA 756  
QY 52 SVPISPVSGRGPDPDAHV 70  
Db 757 RAPEGRPAPE-PTPDGRL 774

RESULT 11  
US-10-036-214-35  
Sequence 35, Application US/10036214  
Publication No. US20030032061A1  
GENERAL INFORMATION:  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
TITLE OF INVENTION: ACTS ENCODING THE SAME  
FILE REFERENCE: P3030R1C11  
CURRENT APPLICATION NUMBER: US/10/036,214  
CURRENT FILING DATE: 2001-12-26  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/112514  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113300  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/113430  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113605  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113621  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/114140  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/115552  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/116843  
PRIOR FILING DATE: 1999-01-22  
PRIOR APPLICATION NUMBER: 60/125774  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 60/125778  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 60/125826  
PRIOR FILING DATE: 1999-03-24  
PRIOR APPLICATION NUMBER: 60/127035  
PRIOR FILING DATE: 1999-03-31  
PRIOR APPLICATION NUMBER: 60/127706  
PRIOR FILING DATE: 1999-04-05  
PRIOR APPLICATION NUMBER: 60/129122

PRIOR FILING DATE: 1999-04-13  
PRIOR APPLICATION NUMBER: 60/130359  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: 60/131270  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131272  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131291  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/132371  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/132379  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/132383  
PRIOR FILING DATE: 1999-05-04  
PRIOR APPLICATION NUMBER: 60/135750  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: 60/138166  
PRIOR FILING DATE: 1999-06-08  
PRIOR APPLICATION NUMBER: 60/144791  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/146970  
PRIOR FILING DATE: 1999-08-03  
PRIOR APPLICATION NUMBER: 60/162506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 09/311832  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 09/380142  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/644848  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 09/747259  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 09/816744  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 09/854208  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: 09/854280  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: 09/874503  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: 09/869599  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: 09/908,827  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: PCT/US99/10733  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30720  
PRIOR FILING DATE: 1999-12-22  
PRIOR APPLICATION NUMBER: PCT/US00/05601  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: PCT/US00/05841  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: PCT/US00/14042  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: PCT/US00/15264  
PRIOR FILING DATE: 2000-06-02  
PRIOR APPLICATION NUMBER: PCT/US00/23522  
PRIOR FILING DATE: 2000-08-23  
PRIOR APPLICATION NUMBER: PCT/US00/23328  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: PCT/US00/32678  
PRIOR FILING DATE: 2000-12-01  
PRIOR APPLICATION NUMBER: PCT/US00/34956  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: PCT/US01/06520  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/17800  
PRIOR FILING DATE: 2001-06-01  
PRIOR APPLICATION NUMBER: PCT/US01/19692  
PRIOR FILING DATE: 2001-06-20



```

: PRIOR APPLICATION NUMBER: PCT/US01/21066
: PRIOR FILING DATE: 2001-06-29
: PRIOR APPLICATION NUMBER: PCT/US01/21735
: PRIOR FILING DATE: 2001-07-09
: NUMBER OF SEQ ID NOS: 80
: SEQ ID NO 35
: LENGTH: 888
: TYPE: PRT
: ORGANISM: Homo Sapien
US-10-036-214-35

Query Match      20.0%; Score 83.5; DB 9; Length 888;
Best Local Similarity 35.4%; Pred. No.0.91;
Matches 28; Conservative 7; Mismatches 33; Indels 11; Gaps 5

OY      1 GTH----SLLRPAAYVPLRMQGPANHVLSFLRP--SWD---LVSAFYSLPLAPLSPT 51
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      698 GRHDLDGSLPTPEQCTPLPQKRLLPTP-HPNHALGPRAWDHGHPLLPASSSLLTLAPA 756
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      52 SVLISPVSVGGRDDPDANHV 70
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      757 RAPEQPPAPGE-PLPDGRL 774
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 12  
US-10-137-865-544  
; Sequence 544, Application US/10137865  
; Publication No. US20030032155A1

```

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Matanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C154
CURRENT APPLICATION NUMBER: US/10/137,865
CURRENT FILING DATE: 2002-05-03
PRIORITY APPLICATION removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 544
LENGTH: 888
TYPE: PRT
ORGANISM: Homo Sapien
US-10-137-865-544

```

Query Match	20.08;	Score 83.5;	DB 9;	Length 888;
Best Local Similarity	35.48;	Pred. No. 0.91;		
Matches 28; Conservative	7;	Mismatches 33;	Indels 11;	Gaps 5

QY 1 GTH---SLPRRAVVPRLKMQGRHNVLSLR-SMD---LVSAFYSLPLAPISPT 51  
| | | | : | | | | : | | | : |  
Db 698 GRHDLDGSLPTEQTRLPGKRLLPTR-NRHNALGRAMDGHNRLLPASASSLLLPARA 756

```

QY      32 SVFISPVSVGKGPDPDAHV 10
          | | : | | | :
Db      757 RAPEQPPAPGE-PTPDGRL 774

```

### RESULT 13

US-10-140-474-544  
; Sequence 544, Application US/10140474  
; Publication No. US20030032156A1  
GENERAL INNOVATIONS

```

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C162
CURRENT APPLICATION NUMBER: US/10/140,474
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 544
LENGTH: 888
TYPE: PRT
ORGANISM: Homo Sapien
US-10-140-474-544

```

Query Match	20.08;	Score 83.5;	DB 9;	Length 888;
Best Local Similarity	35.48;	Pred. No. 0.91;		
Matches 28; Conservative	7;	Mismatches 33;	Indels 11;	Gaps 5

DQ 1 GTH---SLPRRAVVPVLKMQPGRAHVLSLFR-SND---LVSAFYSPLAPLASPT 51  
| | | | : : | | | : ||  
D6 698 GRPDLDSGLTPTEQTPLRQKRLPTR-NPHNLAGRPAMDHGHPRLPASASSLLTLARA 756

```
QY      52  SVPISPVSGRPPDAHV  70
          | | : | | | :
Db      757  RAPEQPPAPGE-PTPDGRL  774
```

RESULT 14

; Sequence 35, Application US/10035719  
; Publication No. US20030036114A1  
; GENERAL INFORMATION:

APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 104

;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/112514

PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113300  
PRIOR FILING DATE: 1998-12-22

PRIOR APPLICATION NUMBER: 60/113430	PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113605	PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621	PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/114140	PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115552	PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/116843	PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/125774	PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125778	PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125826	PRIOR FILING DATE: 1999-03-24
PRIOR APPLICATION NUMBER: 60/127035	PRIOR FILING DATE: 1999-03-31
PRIOR APPLICATION NUMBER: 60/127706	PRIOR FILING DATE: 1999-04-05
PRIOR APPLICATION NUMBER: 60/129122	PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: 60/130359	PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/131270	PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131272	PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131291	PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/132371	PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/132379	PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/132383	PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/135750	PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/138166	PRIOR FILING DATE: 1999-06-08
PRIOR APPLICATION NUMBER: 60/144191	PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/146970	PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/162506	PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 09/311832	PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/380142	PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/654484	PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 09/747259	PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 09/816744	PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/854208	PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/854280	PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/874503	PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: 09/869599	PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: 09/908, 827	PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US99/107333	PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: PCT/US99/28551	PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/307200	

```

? PRIOR FILING DATE: 1999-12-22
? PRIOR APPLICATION NUMBER: PCT/US00/05601
? PRIOR FILING DATE: 2000-03-01
? PRIOR APPLICATION NUMBER: PCT/US00/05841
? PRIOR FILING DATE: 2000-03-02
? PRIOR APPLICATION NUMBER: PCT/US00/14042
? PRIOR FILING DATE: 2000-05-22
? PRIOR APPLICATION NUMBER: PCT/US00/15264
? PRIOR FILING DATE: 2000-06-02
? PRIOR APPLICATION NUMBER: PCT/US00/23552
? PRIOR FILING DATE: 2000-08-23
? PRIOR APPLICATION NUMBER: PCT/US00/23328
? PRIOR FILING DATE: 2000-08-24
? PRIOR APPLICATION NUMBER: PCT/US00/32678
? PRIOR FILING DATE: 2000-12-01
? PRIOR APPLICATION NUMBER: PCT/US00/34956
? PRIOR FILING DATE: 2000-12-20
? PRIOR APPLICATION NUMBER: PCT/US01/06520
? PRIOR FILING DATE: 2001-02-28
? PRIOR APPLICATION NUMBER: PCT/US01/17800
? PRIOR FILING DATE: 2001-06-01
? PRIOR APPLICATION NUMBER: PCT/US01/19692
? PRIOR FILING DATE: 2001-06-20
? PRIOR APPLICATION NUMBER: PCT/US01/21066
? PRIOR FILING DATE: 2001-06-29
? PRIOR APPLICATION NUMBER: PCT/US01/21735
? PRIOR FILING DATE: 2001-07-09
? NUMBER OF SEQ ID NOS: 80
? SEQ ID NO 35
? LENGTH: 888
? TYPE: PRT
? ORGANISM: Homo Sapien
US-10-035-719-35

Query Match          20.0%; Score 83.5; DB 9; Length 888;
Best Local Similarity 35.4%; Pred. No. 0.91;
Matches 28; Conservative 7; Mismatches 33; Indels 11; Gaps 5

QY 1 GTH----SLPRRAVYPLRMQGRPAHVLSFLRP-SWD---LVSAFYSLPLAPLSPT 51
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 698 GPHDLSGLPLRPFGTLPQKRLPFR-HNPHALGPRAMDGHPLPLPASASSLLLLAPA 756
QY 52 SVPISPVYVGKRPDPDAH 70
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 757 RAPEQPPAPGE-PRPDGRL 774

RESULT 15
US-10-142-431-544
; Sequence 544, Application US/10142431
; Publication No. US20030036179a1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvarioff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zhenli
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC251
; CURRENT APPLICATION NUMBER: US/10/142.431

```

```
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 544
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-431-544

Query Match      20.0%; Score 83.5; DB 9; Length 888;
Best Local Similarity 35.4%; Pred. No. 0.91;
Matches 28; Conservative 7; Mismatches 33; Indels 11; Gaps 5;

QY 1 GTH-----SLPRPAAYVPLRMQGPDAHPLYLSFLRP-SWD---LVSATFSLPLAPLSPT 51
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 698 GPHDLDSGLPLPTEQTPLEOKRLPTP-HPHHALGPPAWDHGHPLLPASASSLLILAPA 756
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 52 SVPSISPVSVGRGPPDPDAHV 70
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 757 RAPEQPPAPGE-PTPDGRL 774
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: March 4, 2003, 12:43:43  
Job time : 4.33133 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:25:00 ; Search time 95.9157 Seconds  
(without alignments)  
582.095 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287  
Sequence: 1 MELALCRWGLLLALLPPGA.....YGRGPPDAHVAVNLRYEG 419

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq.101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2181	95.4	419	22	AAE09213 Human p68HER-2 gen
2	2176	95.1	419	22	AAE09212 Human p68HER-2 gen
3	2174	95.1	419	22	AAE09205 Human p68HER-2 gen
4	2172	95.0	419	22	AAE09207 Human p68HER-2 gen
5	2172	95.0	419	22	AAE09209 Human p68HER-2 gen
6	2171	94.9	419	23	AAE20348 Human truncated HE
7	2170	94.9	419	22	AAE09203 Human p68HER-2 gen
8	2169	94.8	419	22	AAE09181 Human p68HER-2 gen
9	2169	94.8	419	22	AAE09208 Human p68HER-2 gen
10	2169	94.8	419	22	AAE09210 Human p68HER-2 gen

11	2168	94.8	419	22	AAE09206 Human p68HER-2 gen
12	2168	94.8	419	22	AAE09211 Human p68HER-2 gen
13	2167	94.8	419	22	AAE09204 Human p68HER-2 gen
14	2164	94.6	419	22	AAE09216 Human p68HER-2 gen
15	2161	94.5	419	22	AAE09200 Human p68HER-2 gen
16	2161	94.5	419	22	AAE09202 Human p68HER-2 gen
17	2158.5	94.4	420	21	AAV97240 Truncated HER-2, p
18	2157	94.3	419	22	AAE09183 Human p68HER-2 gen
19	2155	94.2	419	22	AAE09214 Human p68HER-2 gen
20	2155	94.2	419	22	AAE09215 Human p68HER-2 gen
21	1878	82.1	645	22	AAE09208 Human ErbB2 oncopr
22	1878	82.1	645	22	AAE09209 Human ErbB2 extrac
23	1878	82.1	653	21	AAE21203 Extracellular HER-
24	1878	82.1	653	23	AAE51145 Human HER-2/neu on
25	1878	82.1	712	23	AAE21204 Human HER-2/neu fu
26	1878	82.1	712	23	AAE51149 Her-2/neu extracel
27	1878	82.1	782	18	AAE19764 Her-2-GW-CSF immuno
28	1878	82.1	919	21	AAE21203 Human HER-2/neu pr
29	1878	82.1	919	23	AAE51148 Human HER-2/neu pr
30	1878	82.1	1200	21	AAE21208 Human HER-2/neu pr
31	1878	82.1	1255	17	AAE01111 Human HER-2/neu pr
32	1878	82.1	1255	20	AAE2406 Human HER-2/neu on
33	1878	82.1	1255	21	AAE21198 Human HER-2/neu pr
34	1878	82.1	1255	21	AAE84780 Amino acid sequenc
35	1878	82.1	1255	21	AAE92620 Human heregulin 2
36	1878	82.1	1255	22	AAE12130 Human tyrosine kin
37	1878	82.1	1255	22	AAE85458 Human HER-2/neu pr
38	1878	82.1	1255	22	AAE88267 HER2/neu amino aci
39	1878	82.1	1255	22	AAE60167 HER2 transgene pla
40	1878	82.1	1255	23	AAE24067 Human Her-2 protei
41	1878	82.1	1255	23	AAE20479 Human Her-2/neu pr
42	1878	82.1	1255	23	AAE51143 Human Her-2/neu on
43	1878	82.1	1255	23	AAU77415 Human Her-2/neu po
44	1878	82.1	1255	23	AAU74545 Human HER2 (ErbB2)
45	1860	81.3	1433	14	AAE39568 Sequence of c-erbB

#### ALIGNMENTS

RESULT 1	
AAE09213	
ID	AAE09213 standard; Protein; 419 AA.
XX	
AC	AAE09213;
XX	
DT	15-NOV-2001 (first entry)
XX	
DE	Human p68HER-2 generic protein variant 11.
XX	
KW	HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KW	solid tumour; cancer; polymorphism; cytostatic; gene therapy;
KW	p68HER-2; ECDIIIA; variant.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	1..340
FT	/note= "Identical to N-terminal region of p185HER-2"
FT	341..419
FT	/label= ECDIIIA_variant
FT	/note= "Extracellular domain IIIa variant"
FT	124
FT	/note= "Represented as Agn in the parent sequence shown
FT	in the specification"
FT	125
FT	/note= "Represented as Agn in the parent sequence shown
FT	in the specification"
FT	342
FT	/label= Unknown
FT	345
FT	/label= Unknown
FT	346
FT	Misc-difference

```

FT      /note="p68HER-2 generic sequence (AAE09181) Xaa
FT      substituted with Leu"
FT      Misc-difference 356
FT      /label= Unknown
FT      Misc-difference 358
FT      /label= Unknown
FT      Misc-difference 361
FT      /label= Unknown
FT      Misc-difference 376
FT      /label= Unknown
FT      Misc-difference 394
FT      /label= Unknown
FT      Misc-difference 404
FT      /label= Unknown
FT      Misc-difference 413
FT      /note="p68HER-2 generic sequence (AAE09181) Xaa
FT      substituted with Asn"

```

```

XX      WO200161356-A1.
XX      23-AUG-2001.
XX      16-FEB-2001; 2001WO-US05327.
XX      16-FEB-2000; 2000US-0506079.
XX      (UYOR-) UNIV OREGON HEALTH SCI.
XX      Clinton G, Henner WD, Evans A;
XX      WPI; 2001-529934/58.
XX      New polypeptide, which binds to the extracellular domain of HER-2 for
XX      the treatment of hard tumors -
XX      Example 11; Page -: 61pp; English.

```

```

XX      The invention relates to novel HER-2 (herstatin-2) antagonist
XX      particularly a polypeptide that binds to the extracellular domain (ECD)
XX      of HER-2 at a site that is different from the binding site of humanised
XX      antibody, Herceptin, at an affinity of at least 10-8. The present
XX      invention is based upon the initial discovery of an alternative HER-2
XX      mRNA transcript with 274 bp insert of intron 8. The translation product
XX      of the alternative transcript is a truncated HER-2 protein designated
XX      p68HER-2 which lacks the transmembrane and intracellular domains of
XX      p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIia.
XX      The ECDIIia-containing polypeptides bind tightly to, and thus antagonise
XX      the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
XX      nucleic acids encoding these are useful to treat, diagnose and identify
XX      solid tumours. The present sequence is human p68HER-2 generic protein
XX      containing ECDIIia variant sequence.
XX      Note: The present sequence is not shown in the specification but is
XX      derived from p68HER-2 generic sequence (SEQ ID NO:2) shown in the
XX      sequence listing (AAE09181).

```

XX Sequence 419 AA:

```

Query Match      95.4%; Score 2181; DB 22; Length 419;
Best local Similarity 95.9%; Pred. No. 7, 2e-169;
Matches 402; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```

```

OY      1 METALALCRWGLIALLPAGASTVOCTGTDMLRLPASPEIHLMDLRHLVQGGVQVQNL 60
DB      1 METALALCRWGLIALLPAGASTVOCTGTDCKRLRPASPEIHLMDLRHLVQGGVQVQNL 60
OY      61 EUTLPTNASTSLFDIOEYGVVLAHNOYRQVPLQRLRTIVGTQGFEDNVALAVLDNG 120
DB      61 EUTLPTNASTSLFDIOEYGVVLAHNOYRQVPLQRLRTIVGTQGFEDNVALAVLDNG 120
OY      121 DPLNNTPTVTGASPGSLREIQLSLTEILKGVLIQIRNPOLCYODITLWKDIFHKNNOLA 180
DB      121 DPLNNTPTVTGASPGSLREIQLSLTEILKGVLIQIRNPOLCYODITLWKDIFHKNNOLA 180

```

```

OY      181 LTLIDNRSRACHPCSPMKGSRGSESSDCSLTRTVACAGCARCKPLPTDCCHQC 240
DB      181 LTLIDNRSRACHPCSPCKGSRGSESSDCSLTRTVACAGCARCKPLPTDCCHQC 240
OY      241 AAGCTPRKHSDDLACHFNHSGICEHCPALVYNNDTFESMNPNGRRTFGASCYTACP 300
DB      241 AAGCTPRKHSDDLACHFNHSGICEHCPALVYNNDTFESMNPNGRRTFGASCYTACP 300
OY      301 YNYLSTDVGSCITLVCPLHNOEVAEDGTORCEKSRKARGTSLLRPPAAVPLRMOP 360
DB      301 YNYLSTDVGSCITLVCPLHNOEVAEDGTORCEKSRKARGTSLLRPPAAVPLRMOP 360
OY      361 GPAHPVLSFLRPSWDLVSAFYSLLAPLSPTSVPIPSVSGRCPDDAHVAANLSRYEG 419
DB      361 XPAHPVLSFLRPSWDLVSAFYSLLAPLSPTSVPIPSVSGRCPDDAHVAANLSRYEG 419

```

# RESULT 2

AAE09212 ID AAE09212 standard; Protein; 419 AA.

AC AAE09212;

DT 15-NOV-2001 (first entry)

DE Human p68HER-2 generic protein variant 10.

KW HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;

KW solid tumour; cancer; polymorphism; cytostatic; gene therapy;

KW p68HER-2; ECDIIia; variant.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Region 1..340 "Identical to N-terminal region of p185HER-2"

FT Domain 341..419 "label= ECDIIia-variant"

FT Misc-difference 124 "note= "Extracellular domain Iia variant"

FT Misc-difference 125 "note= "Represented as Agn in the parent sequence shown in the specification"

FT Misc-difference 342 "note= "Represented as Agn in the parent sequence shown in the specification"

FT Misc-difference 345 "label= Unknown"

FT Misc-difference 346 "label= Unknown"

FT Misc-difference 346 "label= Unknown"

FT Misc-difference 356 "label= Unknown"

FT Misc-difference 358 "label= Unknown"

FT Misc-difference 361 "label= Unknown"

FT Misc-difference 376 "label= Unknown"

FT Misc-difference 394 "label= Unknown"

FT Misc-difference 404 "label= Unknown"

FT Misc-difference 413 "note= "p68HER-2 generic sequence (AAE09181) Xaa substituted with Asn"

XX WO200161356-A1.

XX 23-AUG-2001.

XX 16-FEB-2001; 2001WO-US05327.

```

PR 16-FEB-2000; 2000US-0506079.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Clinton G, Henner WD, Evans A;
XX
XX WPI; 2001-529934/58.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
XX the treatment of hard tumors -
XX
XX Example 11; Page -: 61pp; English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
XX particularly a polypeptide that binds to the extracellular domain (ECD)
XX of HER-2 at a site that is different from the binding site of humanised
XX antibody, Hereceptin, at an affinity of at least 108. The present
XX invention is based upon the initial discovery of an alternative HER-2
XX mRNA transcript with 274 bp insert of intron 8. The translation product
XX of the alternative transcript is a truncated HER-2 protein designated
XX p68HER-2 which lacks the transmembrane and intracellular domains of
XX p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
XX The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
XX the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
XX nucleic acids encoding these are useful to treat, diagnose and identify
XX solid tumours. The present sequence is human p68HER-2 generic protein
XX containing ECDIIIA variant sequence.
XX Note: The present sequence is not shown in the specification but is
XX derived from p68HER-2 generic sequence (SEQ ID NO:2) shown in the
XX sequence listing (AAE09181).
XX
XX Sequence 419 AA:
SQ
XX
XX Query Match 95.1%; Score 2176; DB 22; Length 419;
XX Best Local Similarity 95.7%; Pred. No. 1.8e-168;
XX Matches 401; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
XX
XX 1 METALCRWGGLLALLPFAASTOVCTGDMKRLPASPETHLDMRLHYOGCQVVGNN 60
XX 1 METALCRWGGLLALLPFAASTOVCTGDKIRLPASETHLDMRLHYOGCQVVGNN 60
XX
XX 61 ELTYLPTNASLFLDIQIEVGQVYLIAHNOVROVPLQRLRIYNGTOLFEDNVALAVLDNG 120
XX 61 ELTYLPTNASLFLDIQIEVGQVYLCAHNOVROVPLQRLRIYNGTOLFEDNVALAVLDNG 120
XX
XX 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQNRNQLCYQDTILMKDIFHKNNQIA 180
XX 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQNRNQLCYQDTILMKDIFHKNNQIA 180
XX
XX 181 LTLIDNRSRACHPCSPMCKSGESSEDCOSLRTYACAGCARGKGPLPTDCCHEOC 240
XX 181 LTLIDNRSRACHPCSPMCKSGESSEDCOSLRTYACAGCARGKGPLPTDCCHEOC 240
XX
XX 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESMPNPEGRYTFGASCVTACP 300
XX 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESMPNPEGRYTFGASCVTACP 300
XX
XX 301 YNVLSTDVSGCTLVCPCLHNOEYTAEDGTQRCCEKSPCARGTHSLPRAAVPVPLRMQP 360
XX 301 YNVLSTDVSGCTLVCPCLHNOEYTAEDGTQRCCEKSPCARGTHSLPRAAVPVPLRMQP 360
XX
XX 361 GPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVPSVGRCPDDDAHVAVLSRYEG 419
XX 361 GPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVPSVGRCPDDDAHVAVLSRYEG 419
XX
XX RESULT 3
XX AAE09205
XX ID AAE09205 standard; Protein; 419 AA.
XX AC AAE09205;
XX
XX 15-NOV-2001 (first entry)

```

```

XX
XX Human p68HER-2 generic protein variant 3.
DE
XX
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;
XX Solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA; variant.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX Region 1..340
XX /note="Identical to N-terminal region of p185HER-2"
XX Domain 341..419
XX /label="ECDIIIA variant"
XX /note="Extracellular domain IIIA variant"
XX Misc-difference 124
XX /note="Represented as Agn in the parent sequence shown
XX in the specification"
XX Misc-difference 125
XX /note="Represented as Agn in the parent sequence shown
XX in the specification"
XX Misc-difference 342
XX /label="Unknown"
XX Misc-difference 345
XX /label="Unknown"
XX Misc-difference 346
XX /note="p68HER-2 generic sequence (AAE09181) Xaa
XX substituted with Leu"
XX Misc-difference 356
XX /label="Unknown"
XX Misc-difference 358
XX /label="Unknown"
XX Misc-difference 361
XX /label="Unknown"
XX Misc-difference 376
XX /label="Unknown"
XX Misc-difference 394
XX /label="Unknown"
XX Misc-difference 404
XX /label="Unknown"
XX Misc-difference 413
XX /label="Unknown"
XX WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
XX 16-FEB-2000; 2000US-0506079.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Clinton G, Henner WD, Evans A;
XX
XX WPI; 2001-529934/58.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
XX the treatment of hard tumors -
XX
XX Example 11; Page -: 61pp; English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
XX particularly a polypeptide that binds to the extracellular domain (ECD)
XX of HER-2 at a site that is different from the binding site of humanised
XX antibody, Hereceptin, at an affinity of at least 108. The present
XX invention is based upon the initial discovery of an alternative HER-2
XX mRNA transcript with 274 bp insert of intron 8. The translation product
XX of the alternative transcript is a truncated HER-2 protein designated
XX p68HER-2 which lacks the transmembrane and intracellular domains of
XX p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
XX The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
XX the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the

```

CC nucleic acids encoding these are useful to treat, diagnose and identify  
CC solid tumours. The present sequence is human p68HER-2 generic protein  
CC containing ECDIIIA variant sequence.  
CC Note: The present sequence is not shown in the specification but is  
CC derived from p68HER-2 generic sequence (SEQ ID NO:2) shown in the  
CC sequence listing (AAE09181).  
XX  
SO Sequence 419 AA:  
  
Query Match 95.1%; Score 2174; DB 22; Length 419;  
Best Local Similarity 95.7%; Pred. No. 2.7e-168;  
Matches 401; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
  
OY 1 MELAALCRMGILLALLPPGAASVQVCTGDMKRLRASPETHLDMIRHLHYOGCQVVGNL 60  
DB 1 MELAALCRMGILLALLPPGAASVQVCTGDMKRLRASPETHLDMIRHLHYOGCQVVGNL 60  
  
OY 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIYRGTLQFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIYRGTLQFEDNYALAVLDNG 120  
  
OY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180  
  
OY 181 LTLIDTNRSRACHPCSPMKGSRGESSPDCSLTRTVACAGSCARCKGPLPTDCCHEOC 240  
DB 181 LTLIDTNRSRACHPCSPMKGSRGESSPDCSLTRTVACAGSCARCKGPLPTDCCHEOC 240  
  
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRYTFGASCVTACP 300  
  
OY 301 YNTLSTDVSGCTVYCLHNOEYTAEDGTQRCCKSPCARGTSLPRPAAVVPPLRMQP 360  
DB 301 YNTLSTDVSGCTVYCLHNOEYTAEDGTQRCCKSPCARGTSLPRPAAVVPPLRMQP 360  
  
OY 361 GPAHPVLSFLRSPMDLVSAFYSLPLAPLSPTSVISPVSGRGPDPAHVAVLSRYEG 419  
DB 361 GPAHPVLSFLRSPMDLVSAFYSLPLAPLSPTSVISPVSGRGPDPAHVAVLSRYEG 419  
  
RESULT 4  
AAE09207  
ID AAE09207 standard; Protein: 419 AA.  
XX  
AC AAE09207;  
XX  
DT 15-NOV-2001 (first entry)  
XX  
DE Human p68HER-2 generic protein variant 5.  
XX  
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;  
KW solid tumor; cancer; polymorphism; cytostatic; gene therapy;  
KW p68HER-2; ECDIIIA; variant.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FH Region 1..340  
FT /note= "Identical to N-terminal region of p185HER-2"  
FT Domain 341..419  
FT /label= "ECDIIIA variant"  
FT /note= "Extracellular domain IITA variant"  
FT Misc-difference 124  
FT /note= "Represented as Agn in the parent sequence shown  
FT in the specification"  
FT Misc-difference 125  
FT /note= "Represented as Agn in the parent sequence shown  
FT in the specification"  
FT Misc-difference 342  
FT /label= "Unknown"  
FT Misc-difference 345

FT /label= Unknown  
FT Misc-difference 346  
FT /label= Unknown  
FT Misc-difference 356  
FT /label= Unknown  
FT Misc-difference 358  
FT /note= "p68HER-2 generic sequence (AAE09181) Xaa  
FT substituted with Leu"  
FT Misc-difference 361  
FT /label= Unknown  
FT Misc-difference 376  
FT /label= Unknown  
FT Misc-difference 394  
FT /label= Unknown  
FT Misc-difference 404  
FT /label= Unknown  
FT Misc-difference 413  
FT /label= Unknown  
XX  
XX WO200161356-A1.  
XX  
XX 23-AUG-2001.  
XX  
XX 16-FEB-2001; 2001WO-US05327.  
XX  
XX 16-FEB-2000; 2000US-0506079.  
XX  
XX (UYOR-) UNIV OREGON HEALTH SCI.  
XX  
XX Clinton G, Henner WD, Evans A;  
XX  
XX WPI; 2001-529934/58.  
XX  
XX  
XX PT New polypeptide, which binds to the extracellular domain of HER-2 for  
XX the treatment of hard tumors .  
XX  
XX Example 11; Page -: 61pp; English.  
XX  
XX The invention relates to novel HER-2 (herstatin-2) antagonist  
XX particularly a polypeptide that binds to the extracellular domain (ECD)  
XX of HER-2 at a site that is different from the binding site of humanised  
XX antibody, Herceptin, at an affinity of at least 10<sup>-8</sup>. The present  
XX invention is based upon the initial discovery of an alternative HER-2  
XX mRNA transcript with 274 bp insert of intron 8. The translation product  
XX of the alternative transcript is a truncated HER-2 protein designated  
XX p68HER-2 which lacks the transmembrane and intracellular domains of  
XX p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.  
XX The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise  
XX the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the  
XX nucleic acids encoding these are useful to treat, diagnose and identify  
XX solid tumours. The present sequence is human p68HER-2 generic protein  
XX containing ECDIIIA variant sequence.  
XX Note: The present sequence is not shown in the specification but is  
XX derived from p68HER-2 generic sequence (SEQ ID NO:2) shown in the  
XX sequence listing (AAE09181).  
XX  
XX  
XX  
SO Sequence 419 AA:  
  
Query Match 95.0%; Score 2172; DB 22; Length 419;  
Best Local Similarity 95.5%; Pred. No. 3.9e-168;  
Matches 400; Conservative 1; Mismatches 18; Indels 0; Gaps 0;  
  
OY 1 MELAALCRMGILLALLPPGAASVQVCTGDMKRLRASPETHLDMIRHLHYOGCQVVGNL 60  
DB 1 MELAALCRMGILLALLPPGAASVQVCTGDMKRLRASPETHLDMIRHLHYOGCQVVGNL 60  
  
OY 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIYRGTLQFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIYRGTLQFEDNYALAVLDNG 120  
  
OY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180



```
QY 181 LTLIDNRSRACHPCSPCKGSRGCGESSSDCOSLRTTCVACGACRCKGPLPTDCCHEOC 240
    |||
Db 181 LTLIDNRSRACHPCSPCKGSRGCGESSSDCOSLRTTCVACGACRCKGPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCVTACP 300
    |||
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCVTACP 300
QY 301 YNLTSTDVGSCCTLVCPRLHNOEVTAEEDGTORCEKSRPCAGTSHLLPRPAAVPPLRMOP 360
    |||
Db 301 YNLTSTDVGSCCTLVCPRLHNOEVTAEEDGTORCEKSRPCAGTSHLLPRPAAVPPLRMOP 360
QY 361 GPAHPVLSFLRPSWDLVSAFYSLPLAPLSPTSPISPVSGRGPDPDAHVAVNLSTREG 419
    |||
Db 361 XPAHPVLSFLRPSWDLVSAFYSLPLAPLDPSTVXISPSVSGRGKDPDAHVAVNLSTREG 419

RESULT 5
AAE09209 standard; Protein; 419 AA.
XX
AC AAE09209:
XX
DT 15-NOV-2001 (first entry)
XX
DE Human p68HER-2 generic protein variant 7.
XX
KW HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KW solid tumour; cancer; polymorphism; cytostatic; gene therapy;
KW p68HER-2; ECDIIIA; variant.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT Region 1..340
FT /note= "Identical to N-terminal region of p185HER-2"
FT Domain 341..419
FT /label= "ECDIIIA variant"
FT /note= "Extracellular domain IITA variant"
FT Misc-difference 124 /note= "Represented as Agn in the parent sequence shown
    in the specification"
FT Misc-difference 125 /note= "Represented as Agn in the parent sequence shown
    in the specification"
FT Misc-difference 342 /label= Unknown
FT Misc-difference 345 /label= Unknown
FT Misc-difference 346 /label= Unknown
FT Misc-difference 356 /label= Unknown
FT Misc-difference 358 /label= Unknown
FT Misc-difference 361 /label= Unknown
FT Misc-difference 376 /label= Unknown
FT /note= "p68HER-2 generic sequence (AAE09181) Xaa
    substituted with Ile"
FT Misc-difference 394 /label= Unknown
FT Misc-difference 404 /label= Unknown
FT Misc-difference 413 /label= Unknown
FT Misc-difference /label= Unknown
XX
XX WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
PF
```

```
XX
PR 16-FEB-2000; 2000US-0506079.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
PA Clinton G, Henner WD, Evans A;
XX WPI; 2001-529934/58.
XX
PT New polypeptide, which binds to the extracellular domain of HER-2 for
    the treatment of hard tumors -
XX
PS Example 11; Page -: 61pp; English.
XX
CC The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Herceptin, at an affinity of at least 10-8. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is human p68HER-2 generic protein
CC containing ECDIIIA variant sequence.
CC Note: The present sequence is not shown in the specification but is
CC derived from p68HER-2 generic sequence (SBO ID NO:2) shown in the
CC sequence listing (AAE09181).
XX
SQ Sequence 419 AA:
```

```
Query Match 95.0%; Score 2172; DB 22; Length 419;
Best Local Similarity 95.5%; Pred. No. 3,9e-168;
Matches 400; Conservative 1; Mismatches 18; Indels 0; Gaps 0;
```

```
QY 1 MELAALCRWGLLALLPRGAASVOYCTGDMKRLPASPEPTNIDMLRHLYGCGVQVQGNL 60
    |||
Db 1 MELAALCRWGLLALLPRGAASVOYCTGDMKRLPASPEPTNIDMLRHLYGCGVQVQGNL 60
QY 61 ELTYLTNASTLSFLQDIQEVGYVLAHNOVROVPLORLRYVGTOLFEDNVALAVLDNG 120
    |||
Db 61 ELTYLTNASTLSFLQDIQEVGYVLAHNOVROVPLORLRYVGTOLFEDNVALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLREIQRLSLEILKGVLIQRNPQLCYQDTILMKDIFHKNNOLA 180
    |||
Db 121 DPLNNTPTVTGASPGGLREIQRLSLEILKGVLIQRNPQLCYQDTILMKDIFHKNNOLA 180
QY 181 LTLIDNRSRACHPCSPCKGSRGCGESSSDCOSLRTTCVACGACRCKGPLPTDCCHEOC 240
    |||
Db 181 LTLIDNRSRACHPCSPCKGSRGCGESSSDCOSLRTTCVACGACRCKGPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCVTACP 300
    |||
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCVTACP 300
QY 301 YNLTSTDVGSCCTLVCPRLHNOEVTAEEDGTORCEKSRPCAGTSHLLPRPAAVPPLRMOP 360
    |||
Db 301 YNLTSTDVGSCCTLVCPRLHNOEVTAEEDGTORCEKSRPCAGTSHLLPRPAAVPPLRMOP 360
QY 361 GPAHPVLSFLRPSWDLVSAFYSLPLAPLSPTSPISPVSGRGPDPDAHVAVNLSTREG 419
    |||
Db 361 XPAHPVLSFLRPSWDLVSAFYSLPLAPLDPSTVXISPSVSGRGKDPDAHVAVNLSTREG 419
```

```
RESULT 6
AAE20348
ID AAE20348 standard; Protein; 419 AA.
XX
XX AAE20348;
XX
```

DT 18-JUN-2002 (first entry)  
XX Human truncated HER2 protein.  
XX Human; tumour; endothelial growth factor receptor; EGFR; cytostatic;  
KW herstatin; HER-2 receptor tyrosine kinase; squamous cell carcinoma; lung;  
XX colon; glial cell tumour; cell growth.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 342 /Label= Thr, Ser  
FT Misc-difference 345 /Label= Leu, Pro  
FT Misc-difference 346 /Label= Pro, Leu  
FT Misc-difference 356 /Label= Leu, Gln  
FT Misc-difference 358 /Label= Met, Leu  
FT Misc-difference 361 /Label= Gly, Asp, Ala, Val  
FT Misc-difference 376 /Label= Leu, Ile  
FT Misc-difference 394 /Label= Pro, Arg  
FT Misc-difference 404 /Label= Pro, Leu  
FT Misc-difference 413 /Label= Asp, Asn  
XX  
XX WO200214470-A2.  
XX  
XX 21-FEB-2002.  
XX  
XX 14-AUG-2001; 2001WO-US25502.  
XX  
XX 14-AUG-2000; 2000US-0638834.  
XX  
XX (UYOR-) UNIV OREGON HEALTH SCI.  
XX  
XX Clinton GM;  
XX  
XX WPI; 2002-269185/31.  
XX  
XX  
XX Treating solid tumor characterized by expression of endothelial growth  
PT factor receptor, involves administering recombinant herstatin that  
PT binds to extracellular domain of the endothelial growth factor receptor  
XX  
XX  
PS Claim 1; Page 78-80; 82pp; English.  
XX  
XX The present invention relates to a method for treating a solid tumour  
CC characterised by endothelial growth factor receptor (EGFR) expression.  
CC The method involves administering an agent that binds to an extracellular  
CC domain (ECD) of EGFR. The invention also relates to a naturally occurring  
CC inhibitor of HER-2 receptor tyrosine kinase called herstatin. The co-  
CC expression of herstatin with p185HER2 causes a striking reduction in cell  
CC growth that corresponds with suppression of p185 autophosphorylation. The  
CC method or a pharmaceutical composition is useful for treating a solid  
CC tumour (selected from squamous cell carcinoma, lung carcinoma, colon  
CC carcinoma and glial cell tumour) characterised by EGFR expression. The  
CC present sequence is human truncated HER2 protein that lacks transmembrane  
CC and intracellular domains.  
XX  
XX Sequence 419 AA;  
XX  
XX Query Match 94.9%; Score 2171; DB 23; Length 419;  
XX Best Local Similarity 95.5%; Pred. No. 4.7e-168;  
XX Matches 400; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 1 METALCRMGILLALLPPGAASOVCTGTGDKLRPLPASPETHLMDLRHLHYGCGVVOGNTL 60  
QY 61 ELTYLPTNASTSLFDIOIEVOGVILIAHNOVQVPLQRLRIYRGQLFEDNVALAVLDNG 120  
Db 61 ELTYLPTNASTSLFDIOIEVOGVILCAHNOVQVPLQRLRIYRGQLFEDNVALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGSLREQLRSLEILKGGVLIQORNQOLCYQDTILMKDIFHKNNOLA 180  
Db 121 DPLRRTPTVTGASPGSLREQLRSLEILKGGVLIQORNQOLCYQDTILMKDIFHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKGSRKWESESDCQSLRTVCAGGCARCKPLPTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSRKWESESDCQSLRTVCAGGCARCKPLPTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRTTFGASCTYAC 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRTTFGASCTYAC 300  
QY 301 YNLTSDVGSCTVCPPLHNOEVTAEEDGTORCEKSPCARGTSHLLPPRAAVPPLMKP 360  
Db 301 YNLTSDVGSCTVCPPLHNOEVTAEEDGTORCEKSPCARGTSHLLPPRAAVPPLMKP 360  
QY 361 GPAHPVLSFLRPSWDLVSAFYSLLAPLSPTSVSPISPVSGRGPDDAHVAVNLSRYEG 419  
Db 361 XPAHPVLSFLRPSWDLVSAFYSLLAPLSPTSVSPISPVSGRGPDDAHVAVNLSRYEG 419  
RESULT 7  
AAE09203  
ID AAE09203 standard; Protein; 419 AA.  
XX  
XX AAE09203;  
XX  
XX 15-NOV-2001 (first entry)  
XX  
XX Human p68HER-2 generic protein variant 1.  
XX  
XX  
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;  
KW solid tumour; cancer; polymorphism; cytostatic; gene therapy;  
XX p68HER-2; ECDIIIA; variant.  
XX  
XX Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Region 1..340  
FT Domain /note= "Identical to N-terminal region of p185HER-2"  
FT /label= ECDIIIA\_variant  
FT /note= "Extracellular domain IIA variant"  
FT Misc-difference 124  
FT /note= "Represented as Agn in the parent sequence shown  
FT in the specification"  
FT Misc-difference 125  
FT /note= "Represented as Agn in the parent sequence shown  
FT in the specification"  
FT Misc-difference 342  
FT /note= "p68HER-2 generic sequence (AAE09181) Xaa  
FT substituted with Ser"  
FT Misc-difference 345  
FT /label= Unknown  
FT Misc-difference 346  
FT /label= Unknown  
FT Misc-difference 356  
FT /label= Unknown  
FT Misc-difference 358  
FT /label= Unknown  
FT Misc-difference 361  
FT /label= Unknown  
FT Misc-difference 376  
FT /label= Unknown  
FT Misc-difference 394  
FT /label= Unknown

```

FT Misc-difference 404
FT /label= Unknown
FT Misc-difference 413
FT /label= Unknown
XX
XX WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
XX 16-FEB-2000; 2000US-0506079.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Clinton G, Henner WD, Evans A;
XX
XX WPI; 2001-529934/58.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
XX the treatment of hard tumors -
XX
XX Example 11; Page -: 61pp; English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
XX particularly a polypeptide that binds to the extracellular domain (ECD)
XX of HER-2 at a site that is different from the binding site of humanised
XX antibody, Hereceptin, at an affinity of at least 10-8. The present
XX invention is based upon the initial discovery of an alternative HER-2
XX mRNA transcript with 274 bp insert of intron 8. The translation product
XX of the alternative transcript is a truncated HER-2 protein designated
XX p68HER-2 which lacks the transmembrane and intracellular domains of
XX p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
XX The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
XX the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
XX nucleic acids encoding these are useful to treat, diagnose and identify
XX solid tumours. The present sequence is human p68HER-2 generic protein
XX containing ECDIIIA variant sequence.
XX Note: The present sequence is not shown in the specification but is
XX derived from p68HER-2 generic sequence (SEQ ID NO:2) shown in the
XX sequence listing (AAE09181).
XX
XX Sequence 419 AA:
XX
XX Query Match 94.9%; Score 2170; DB 22; Length 419;
XX Best Local Similarity 95.5%; Pred No. 5.7e-168;
XX Matches 400; Conservative 1; Mismatches 18; Indels 0; Gaps 0;

```

```

DB 361 XPAHPVLSFLRPSWMDVSAFYSLPLAPLPTSVISPVSGRQDPDAHVAVNLSRYEG 419
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
RESULT 8
ID AAE09181 standard; Protein; 419 AA.
XX
XX AAE09181;
AC
XX 15-NOV-2001 (first entry)
XX
XX Human p68HER-2 generic sequence #1.
XX
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;
XX solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 1..340
XX FT /note= "Identical to N-terminal region of p185HER-2"
XX FT Domain 341..419
XX FT /label= ECDIIIA
XX FT /note= "Extracellular domain IIIA"
XX FT Misc-difference 124 /note= "Represented as Agn in the sequence shown in
XX the specification"
XX FT Misc-difference 125 /note= "Represented as Agn in the sequence shown in
XX the specification"
XX FT Misc-difference 342 /label= Unknown
XX FT /note= "Encoded by WCC"
XX FT Misc-difference 345 /label= Unknown
XX FT /note= "Encoded by CYG"
XX FT Misc-difference 346 /label= Unknown
XX FT /note= "Encoded by CYC"
XX FT Misc-difference 356 /label= Unknown
XX FT /note= "Encoded by CYC"
XX FT Misc-difference 358 /label= Unknown
XX FT /note= "Encoded by AGC"
XX FT Misc-difference 361 /label= Unknown
XX FT /note= "Encoded by GNC"
XX FT Misc-difference 376 /label= Unknown
XX FT /note= "Encoded by WTA"
XX FT Misc-difference 389 /note= "Encoded by AGC"
XX FT Misc-difference 394 /label= Unknown
XX FT /note= "Encoded by CST"
XX FT Misc-difference 404 /label= Unknown
XX FT /note= "Encoded by CYG"
XX FT Misc-difference 413 /label= Unknown
XX FT /note= "Encoded by SAC"
XX
XX WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
XX 16-FEB-2000; 2000US-0506079.
XX
XX
XX

```

Query	Subject	Score	DB	Length	419
Query Match	94.8%	Score 2169	DB 22	Length 419	
Best Local Similarity	95.5%	Pred. No. 6, 8e-168			
Matches 400	Conservative 0	Mismatches 19	Indels 0	Gaps 0	
1	MELALACRMGLLALLPFGAASOVCTGTDMLRLPASPETHLMDLRHLYOGQVQVQGNL	60			
1	MELALACRMGLLALLPFGAASOVCTGTDMLRLPASPETHLMDLRHLYOGQVQVQGNL	60			
61	ELTYLPTNASLSFLQDIOEVQGYVLIANQVROVPLQRLRIYRGTOLEFEDNALAVLNG	120			
61	ELTYLPTNASLSFLQDIOEVQGYVLIANQVROVPLQRLRIYRGTOLEFEDNALAVLNG	120			
61	ELTYLPTNASLSFLQDIOEVQGYVLIANQVROVPLQRLRIYRGTOLEFEDNALAVLNG	120			
121	DLNNPTPTVPGASPGSLRELQRLSLLEILKGGVLIQNRNOLCYODITLWKDIFHKNQOLA	180			
121	DLNNPTPTVPGASPGSLRELQRLSLLEILKGGVLIQNRNOLCYODITLWKDIFHKNQOLA	180			
121	DLNNPTPTVPGASPGSLRELQRLSLLEILKGGVLIQNRNOLCYODITLWKDIFHKNQOLA	180			
181	LTLIDITNRSRACHPCSPMKGSKRCWESSESDCSILRTVACAGACARCKPLPTDCCHQC	240			
181	LTLIDITNRSRACHPCSPMKGSKRCWESSESDCSILRTVACAGACARCKPLPTDCCHQC	240			
241	AAGCTGPKHSDDLCLHFNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACR	300			
241	AAGCTGPKHSDDLCLHFNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACR	300			
301	YVYLSITDVQSSCLVCPDLHNOEYTAEGTCQRCCKSPCARGTHSLPLPRAAVVPLRMOP	360			
301	YVYLSITDVQSSCLVCPDLHNOEYTAEGTCQRCCKSPCARGTHSLPLPRAAVVPLRMOP	360			
301	YVYLSITDVQSSCLVCPDLHNOEYTAEGTCQRCCKSPCARGTHSLPLPRAAVVPLRMOP	360			
361	GPANHVLSFLRSWMDLVSAFSLPLAPLPTSPVPIPSVVGRCRDPDAVAANVLSYEG	419			
361	GPANHVLSFLRSWMDLVSAFSLPLAPLPTSPVPIPSVVGRCRDPDAVAANVLSYEG	419			
361	GPANHVLSFLRSWMDLVSAFSLPLAPLPTSPVPIPSVVGRCRDPDAVAANVLSYEG	419			

KW	solid tumour; cancer; polymorphism; cytostatic; gene therapy;	
KM	p68HER-2; ECDIIIA; variant.	
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Region	1..340
FT		/note= "Identical to N-terminal region of p185HER-2"
FT	Domain	341..419
FT		/label= ECDIIIA.variant
FT		/note= "Extracellular domain IIIa variant"
FT	Misc-difference	124
FT		/note= "Represented as Agn in the parent sequence shown in the specification"
FT	Misc-difference	125
FT		/note= "Represented as Agn in the parent sequence shown in the specification"
FT	Misc-difference	342
FT		/label= Unknown
FT	Misc-difference	345
FT		/label= Unknown
FT	Misc-difference	346
FT		/label= Unknown
FT	Misc-difference	356
FT		/label= Unknown
FT	Misc-difference	358
FT		/label= Unknown
FT	Misc-difference	361
FT		/label= Asp, Ala, Val
FT	Misc-difference	376
FT		/label= Unknown
FT	Misc-difference	394
FT		/label= Unknown
FT	Misc-difference	404
FT		/label= Unknown
FT	Misc-difference	413
FT		/label= Unknown
XX		
PM	WO200161356-A1.	
XX		
PD	23-AUG-2001.	
XX		
PF	16-FEB-2001; 2001WO-US05327.	
XX		
PR	16-FEB-2000; 2000US-0506079.	
XX		
PA	(UYOR-) UNIV OREGON HEALTH SCI.	
XX		
PI	Clinton G, Henner WD, Evans A;	
XX		
DR	WPI; 2001-52934/58.	
XX		
PT	New polypeptide, which binds to the extracellular domain of HER-2 for the treatment of hard tumors -	
XX		
PS	Example 11: Page -: 61pp; English.	
XX		
CC	The invention relates to novel HBR-2 (herstatin-2) antagonist particularly a polypeptide that binds to the extracellular domain (ECD) of HER-2 at a site that is different from the binding site of humanised antibody, Herceptin, at an affinity of at least 10 <sup>7</sup> 8. The present invention is based upon the initial discovery of an alternative HER-2 mRNA transcript with 274 bp insert of intron 8. The translation product of the alternative transcript is a truncated HER-2 protein designated p68HER-2 which lacks the transmembrane and intracellular domains of p185HER-2 but contains ECD 1, II of the p185HER-2 and the novel ECDIIIA. The ECDIIIA-containing polypeptides bind tightly to, and thus antagonists the HER-2 receptor. The peptides, which bind to an HBR-2 ECD, and the nucleic acids encoding these are useful to treat, diagnose and identify solid tumours. The present sequence is human p68HER-2 generic protein containing ECDIIIA variant sequence.	
CC	Note: The present sequence is not shown in the specification but is derived from p68HER-2 generic sequence (SEQ ID NO:2) shown in the	

```

CC sequence listing (AAE09181).
XX
SQ Sequence 419 AA:
Query Match 94.8%; Score 2169; DB 22; Length 419;
Best Local Similarity 95.5%; Pred. No. 6.8e-168;
Matches 400; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 1 MELAALCRWGLLALLPAGASTOVCTGDMKRLPASPEHIDMLRHLYOGCQVVGNL 60
DB 1 MELAALCRWGLLALLPAGASTOVCTGDMKRLPASPEHIDMLRHLYOGCQVVGNL 60
OY 61 ELTYLPTNASLSFLDIOEVQGVLLAHNOVQVPLQRLRIYRGTOLEFDNVALAVLNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGVLLAHNOVQVPLQRLRIYRGTOLEFDNVALAVLNG 120
OY 121 DPLNNTTPTVTGASPGGLRELOLRSLTEILKGVLIORNPOLCYQDTILMKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTGASPGGLRELOLRSLTEILKGVLIORNPOLCYQDTILMKDIFHKNNOLA 180
OY 181 LRLIDNRSRACHPCSPCKGSRWGESSEDCOSLTRVACAGCARKGKPLPTDCCHQC 240
DB 181 LRLIDNRSRACHPCSPCKGSRWGESSEDCOSLTRVACAGCARKGKPLPTDCCHQC 240
OY 241 AAGCTGPKHSDCLAHFNHSGICELCPALVYNTDFESMPNPGRYTFGASCTACP 300
DB 241 AAGCTGPKHSDCLAHFNHSGICELCPALVYNTDFESMPNPGRYTFGASCTACP 300
OY 301 YNLTSDVSCSLVCPPLHNOEYTAEDGTORCEKSKPCARGTHSLPRPAVPPLRMOP 360
DB 301 YNLTSDVSCSLVCPPLHNOEYTAEDGTORCEKSKPCARGTHSLPRPAVPPLRMOP 360
OY 361 GFAHPVLSFLRSPMDVSAFYSPLAPLSPTSVPISPVGREGPDDAHVAVALSYEG 419
DB 361 GFAHPVLSFLRSPMDVSAFYSPLAPLSPTSVPISPVGREGPDDAHVAVALSYEG 419

RESULT 10
AAE09210
ID AAE09210 standard; Protein; 419 AA.
XX
AC AAE09210;
XX
DE 15-NOV-2001 (first entry)
XX
DE Human p68HER-2 generic protein variant 8.
XX
KW HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KW solid tumor; cancer; polymorphism; cytostatic; gene therapy;
KW p68HER-2; ECDIIIA; variant.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Region 1..340
FT /note= "Identical to N-terminal region of p185HER-2"
FT Domain 341..419
FT /label= ECDIIIA_variant
FT /note= "Extracellular domain IIA variant"
FT Misc-difference 124
FT /note= "Represented as Agn in the parent sequence shown
FT in the specification"
FT Misc-difference 125
FT /note= "Represented as Agn in the parent sequence shown
FT in the specification"
FT Misc-difference 342
FT /label= Unknown
FT Misc-difference 345
FT /label= Unknown
FT Misc-difference 346
FT /label= Unknown
FT Misc-difference 356
FT /label= Unknown

```

```

FT Misc-difference 358
FT /label= Unknown
FT Misc-difference 361
FT /label= Unknown
FT Misc-difference 376
FT /label= Unknown
FT Misc-difference 394
FT /note= "p68HER-2 generic sequence (AAE09181) Xaa
FT substituted with Arg"
FT Misc-difference 404
FT /label= Unknown
FT Misc-difference 413
FT /label= Unknown
PN MO200161356-A1.
XX
PD 23-AUG-2001.
XX
PF 16-FEB-2001; 2001WO-US05327.
XX
PR 16-FEB-2000; 2000US-0506079.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Clinton G, Henner WD, Evans A;
XX
PI WPI; 2001-529934/58.
XX
PT New polypeptide, which binds to the extracellular domain of HER-2 for
PT the treatment of hard tumors -
XX
PS Example 11; Page -: 61pp: English.
XX
CC The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Herceptin, at an affinity of at least 10-8. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is human p68HER-2 generic protein
CC containing ECDIIIA variant sequence.
CC Note: The present sequence is not shown in the specification but is
CC derived from p68HER-2 generic sequence (SEQ ID NO:2) shown in the
CC sequence listing (AAE09181).
XX
SQ Sequence 419 AA:
Query Match 94.8%; Score 2169; DB 22; Length 419;
Best Local Similarity 95.5%; Pred. No. 6.8e-168;
Matches 400; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 1 MELAALCRWGLLALLPAGASTOVCTGDMKRLPASPEHIDMLRHLYOGCQVVGNL 60
DB 1 MELAALCRWGLLALLPAGASTOVCTGDMKRLPASPEHIDMLRHLYOGCQVVGNL 60
OY 61 ELTYLPTNASLSFLDIOEVQGVLLAHNOVQVPLQRLRIYRGTOLEFDNVALAVLNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGVLLAHNOVQVPLQRLRIYRGTOLEFDNVALAVLNG 120
OY 121 DPLNNTTPTVTGASPGGLRELOLRSLTEILKGVLIORNPOLCYQDTILMKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTGASPGGLRELOLRSLTEILKGVLIORNPOLCYQDTILMKDIFHKNNOLA 180
OY 181 LRLIDNRSRACHPCSPCKGSRWGESSEDCOSLTRVACAGCARKGKPLPTDCCHQC 240
DB 181 LRLIDNRSRACHPCSPCKGSRWGESSEDCOSLTRVACAGCARKGKPLPTDCCHQC 240

```

```

QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTFGASCVTACP 300
   |||
DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTFGASCVTACP 300
QY 301 YNLTSDVSGCTLVCPDLHQEVTAEEDGTORCEKSPCARGTHSLPRPAAVVPRLMOP 360
   |||
DB 301 YNLTSDVSGCTLVCPDLHQEVTAEEDGTORCEKSPCARGTHSLPRPAAVVPRLMOP 360
QY 361 GPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPVSPVSGRGPDDAHVAVNLSTYEG 419
   |||
DB 361 XPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPVSPVSGRGPDDAHVAVNLSTYEG 419

RESULT 11
AAE09206
ID AAE09206 standard: Protein; 419 AA.
AC AAE09206;
DE 15-NOV-2001 (first entry)
XX Human p68HER-2 generic protein variant 4.
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;
XX solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA; variant.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1..340
FT /note= "Identical to N-terminal region of p185HER-2"
FT Domain 341..419
FT /label= ECDIIIA variant
FT /note= "Extracellular domain IIIa variant"
FT Misc-difference 124 /note= "Represented as Agn in the parent sequence shown
   in the specification"
FT Misc-difference 125 /note= "Represented as Agn in the parent sequence shown
   in the specification"
FT Misc-difference 342 /label= Unknown
FT Misc-difference 345 /label= Unknown
FT Misc-difference 346 /label= Unknown
FT Misc-difference 356 /label= Unknown
FT /note= "p68HER-2 generic sequence (AAE09181) Xaa
   substituted with Gln"
FT Misc-difference 358
FT /label= Unknown
FT Misc-difference 361
FT /label= Unknown
FT Misc-difference 376 /label= Unknown
FT Misc-difference 394 /label= Unknown
FT Misc-difference 404 /label= Unknown
FT Misc-difference 413 /label= Unknown
FT Misc-difference /label= Unknown
FT
FT WO200161356-A1.
FT
FT 23-AUG-2001.
FT
FT 16-FEB-2001; 2001WO-0505327.
FT
FT 16-FEB-2000; 2000US-0506079.
FT
FT (UOR-) UNIV OREGON HEALTH SCI.
FT
XX
XX

```

```

PI Clinton G, Henner WD, Evans A;
XX
DR WPI: 2001-529934/58.
XX
PT New polypeptide, which binds to the extracellular domain of HER-2 for
   the treatment of hard tumors.
XX
PS Example 11; Page -: 61pp; English.
XX
CC The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Hereceptin, at an affinity of at least 10-8. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is human p68HER-2 generic protein
CC containing ECDIIIA variant sequence.
CC Note: The present sequence is not shown in the specification but is
CC derived from p68HER-2 generic sequence (SEQ ID NO:2) shown in the
CC sequence listing (AAE09181).
XX
SQ Sequence 419 AA:
XX
Query Match 94.8%; Score 2168; DB 22; Length 419;
Best Local Similarity 95.5%; Pred. No. 8,2e-168;
Matches 400; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 1 MELAALCRMGILLALLPRAASTVOCTGDMKRLRPASPTFLDMIRHLXGCGVVGML 60
   |||
DB 1 MELAALCRMGILLALLPRAASTVOCTGDMKRLRPASPTFLDMIRHLXGCGVVGML 60
QY 61 ELTYLPTNASTSLFLQDQEVQGVYLAHNOVQVPLQRLRIYRGQLFEDNALAVLDNG 120
   |||
DB 61 ELTYLPTNASTSLFLQDQEVQGVYLAHNOVQVPLQRLRIYRGQLFEDNALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSTELKGGVLIQRPOLCYODTILMKDIFHKNNOLA 180
   |||
DB 121 DPLNNTPTVTGASPGGLRELQRLSTELKGGVLIQRPOLCYODTILMKDIFHKNNOLA 180
QY 181 LTLIDTNRSRACHPCSPCKGSKCKWGESESDOSLRTVCAGGACRKPRLPTDCHEOC 240
   |||
DB 181 LTLIDTNRSRACHPCSPCKGSKCKWGESESDOSLRTVCAGGACRKPRLPTDCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTFGASCVTACP 300
   |||
DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTFGASCVTACP 300
QY 301 YNLTSDVSGCTLVCPDLHQEVTAEEDGTORCEKSPCARGTHSLPRPAAVVPRLMOP 360
   |||
DB 301 YNLTSDVSGCTLVCPDLHQEVTAEEDGTORCEKSPCARGTHSLPRPAAVVPRLMOP 360
QY 361 GPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPVSPVSGRGPDDAHVAVNLSTYEG 419
   |||
DB 361 XPAHPVLSFLRPSMDVSAFYSLPLAPLSPTSPVSPVSGRGPDDAHVAVNLSTYEG 419

RESULT 12
AAE09211
ID AAE09211 standard: Protein; 419 AA.
AC AAE09211;
DE 15-NOV-2001 (first entry)
XX Human p68HER-2 generic protein variant 9.
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Hereceptin;
XX
XX

```

KM		Solid tumour: cancer; polymorphism; cytostatic; gene therapy;
KW		p68HER-2; ECDIIIA; variant.
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Region	1..340
FT		/note= "Identical to N-terminal region of p185HER-2"
FT	Domain	341..419
FT		/label= ECDIIIA.variant
FT		/note= "Extracellular domain IIIa variant"
FT	Misc-difference	124
FT		/note= "Represented as Agn in the parent sequence shown in the specification"
FT	Misc-difference	125
FT		/note= "Represented as Agn in the parent sequence shown in the specification"
FT	Misc-difference	342
FT		/label= Unknown
FT	Misc-difference	345
FT		/label= Unknown
FT	Misc-difference	346
FT		/label= Unknown
FT	Misc-difference	356
FT		/label= Unknown
FT	Misc-difference	358
FT		/label= Unknown
FT	Misc-difference	361
FT		/label= Unknown
FT	Misc-difference	376
FT		/label= Unknown
FT	Misc-difference	394
FT		/label= Unknown
FT	Misc-difference	404
FT		/note= "p68HER-2 generic sequence (AAE03181) Xaa substituted with Leu"
FT	Misc-difference	413
FT		/label= Unknown
XX		
PM	WO200161356-A1.	
XX		
PD	23-AUG-2001.	
XX		
PF	16-FEB-2001; 2001WO-US05327.	
XX		
PR	16-FEB-2000; 2000US-0506079.	
XX		
PA	(UYOR-) UNIV OREGON HEALTH SCI.	
XX		
PI	Clinton G, Henner WD, Evans A;	
XX		
DR	WPI: 2001-529934/58.	
XX		
PT	New polypeptide, which binds to the extracellular domain of HER-2 for the treatment of hard tumors -	
XX		
PS	Example 11; Page -: 61pp; English.	
XX		
CC	The invention relates to novel HER-2 (herstatin-2) antagonist particularly a polypeptide that binds to the extracellular domain (ECD) of HER-2 at a site that is different from the binding site of humanised antibody, Herceptin, at an affinity of at least 10 <sup>7</sup> 8. The present invention is based upon the initial discovery of an alternative HER-2 mRNA transcript with 274 bp insert of intron 8. The translation product of the alternative transcript is a truncated HER-2 protein designated p68HER-2 which lacks the transmembrane and intracellular domains of p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA. The ECDIIIA-containing polypeptides bind tightly to, and thus antagonists the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the nucleic acids encoding these are useful to treat, diagnose and identify solid tumours. The present sequence is human p68HER-2 generic protein containing ECDIIIA variant sequence.	
CC	Note: The present sequence is not shown in the specification but is	

CC	derived from p68HER-2 generic sequence (SEQ ID NO:2) shown in the
CC	sequence listing (AAE09181).
XX	
XX	
SO	Sequence 419 AA:
	Query Match 94.8%; Score 2168; DB 22; Length 419;
	Best Local Similarity 95.5%; Pred. No. 8.2e-168;
	Matches 400; Conservative 0; Mismatches 19; Indels 0; Gaps
OY	1 MELALRCRWGLLALLPRGASNOVOTGDMLRLPASPENHDMLRHLVQGCQVVGNL 60
DB	1 MELALRCRWGLLALLPRGASNOVOTGDCKLRLPASPEHMDLMRLLXYGCCQVVGNL 60
OY	61 ELTYLPTNALSFLDIIOEVGYVLIAHQVQVPLQLRIATVRGTOLFEDNYALAVLDNG 120
DB	61 ELTYLPTNALSFLDIDQIEGVYVLCANHQVQVPLQLRIATVRGTOLFEDNYALAVLDNG 120
OY	121 DPLNNTPVTGASPGGLRELQLRSLEILKGSVLIQRNPQLCYDTIIMKDIFHKNNQLA 180
DB	121 DPLXSTPYVGASPGGRRELOLSLETECLKGSVLIQRNPQLCYDTIIMKDIFHKNNQLA 180
OY	181 LTLIDTNRSRACHPCSPMKCGSNCKWSESSEDQSITRVCAAGCARCKGPLPTDCHEQC 240
DB	181 LTLIDTNRSRACHPCSPCKCGSKCKWSESSEDDQSILTRVCAGGACARCKGPLPTDCCHEQC 240
OY	241 AAGCTGKRHSDDCIACLFHNHSGICELCHCPALVTYTDFEEMPMPEGRTFGASCVTACP 300
DB	241 AAGCTGKRHSDDCIACLFHNHSGICELCHCPALVTYTDFESCPRNEGRTYTFGASCVTACP 300
OY	301 YNVLSTDVSGCTLVCPLNHOEVTAEDGTORCEKCSKPCARGTHSLRPAPAVPYLRMPQ 360
DB	301 YNKLTSDVSGCTLVCPNLHNOEVTAEDGTORCEKCSKPCARGHXXP RPAPAVPYLRXPQ 360
OY	361 GPAPPVLSFLRPSWDLYSAFYSLPLAPISPTSPVSIVSGRGDPDAHAVALVLSREG 419
DB	361 XPAPPVLSFLRPSMDYSAFYSLPLAPIDPTSVXISPVSIGRGDPDAHAVALVLSRYEG 419
	RESULT_13
AAE09204	
ID	AAE09204 standard; Protein: 419 AA.
XX	
XX	AAE09204;
XX	
DE	15-NOV-2001 (first entry)
XX	
DE	Human p68HER-2 generic protein variant 2.
XX	
KM	HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KM	solid tumour; cancer; polymorphism; cytostatic; gene therapy;
KM	p68HER-2; ECDIIia; variant.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Region 1..340
FT	/note= "Identical to N-terminal region of p185HER-2"
FT	Domain 341...419
FT	/label= "ECDIIia_variant"
FT	/note= "Extracellular domain IIia variant"
FT	Misc-difference 124
FT	/note= "Represented as Agn in the parent sequence shown
FT	in the specification"
FT	Misc-difference 125
FT	/note= "Represented as Agn in the parent sequence shown
FT	in the specification"
FT	Misc-difference 342
FT	/label= Unknown
FT	Misc-difference 345
FT	/note= "p68HER-2 generic sequence (AAE09181) Xaa
FT	substituted with Pro"
FT	Misc-difference 346
FT	/label= Unknown







CC p18HER-2 but contains ECD 1, II of the p18HER-2 and the novel ECDIIIa  
CC p18HER-2 containing polypeptides blind tightly to, and thus antagonise  
CC the ECDIIa-containing polypeptides bind tightly to, and thus antagonise  
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the  
CC nucleic acids encoding these are useful to treat, diagnose and identify  
CC solid tumours. The present sequence is human p68HER-2 genetic protein  
CC containing ECDIIa variant sequence.  
CC Note: The present sequence is not shown in the specification but is  
CC derived from p68HER-2 genetic sequence (SEQ ID NO:2) shown in the  
CC sequence listing (AA609181).

**SQ** Sequence 419 AA;

Query	Match
Post 1000	

94.58; Score 2161; DB 22; Length 419;

```
Best Local Similarity 95.26; Pred. NO. 36-16/17
Matches 399; Conservative 0; Mismatches 2
```

Matches	399;	Conservative	0;	Mismatches	20;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	-----	--------	----	------	----

QY	1	MELALACWGLLLALPRGASASTOVGCTDKLKLPA	SPETHIDMLHEHVGCGOVGNL	60
Db	1	MELALACWGLLLALPRGASASTOVCTGCTDKLKLPA	SPETHIDMLHNLVGGCGOVGNL	60
QY	61	ELTYLPTNASLSFLQIDQIOWGVYVLI	AHNOYRQVPLRLRIVRGSTQLEEDNVALAYLVDNG	120
Db	61	ELTYLPTNASLSFLQIDQIOWGVYVLI	CAHNOYRQVPLRLRIVRGSTQLEEDNVALAYLVDNG	120
QY	121	DPLNNTPTVATASFGGURELELDSLT	ELKGVLIQRNPOLCTODTILMKDIFHKNNQLA	180
Db	121	DPLXXTPPTVATASFGGURELELDSLT	ELCKLKGVLIQNNPOLCYODTILAMDIFHKNNQLA	180
QY	181	LTLIDTNSRACHPGSPCKGSR	CGWGESSEDDCGSLTFETVACGACRKGRLPDDC	240
Db	181	LTLIDTNSRACHPGSPCKGSR	CGWGESSEDDCGSLTLTVACGACRKGRLPDDC	240
QY	241	AAGCTGPRKSHDCLALCFH	NHSGICELHCPALVYNTDTFESMPNDEGRYTFGASCVTACP	300
Db	241	AAGCTGPRKSHDCLALCFH	NHSGICELHCPALVYNTDTFESCPNDEGRYTFGASCVTACP	300
QY	301	YNYLSTDVGSCITLVCPL	HNDEVTAEDETQRCCKSCPCACGTHSLLPRAPAVPVPLRMOP	360
Db	301	YNYLSTDVGSCITLVCPL	HNDEVTAEDETQRCCKSCPCAGHSHXPRPAVPVPCXOP	360
QY	361	GPAPHVLSFLARPMSMDIVSA	FSFLAPLPLSPISVYISVYSGRGEPDPAHVAVNL	419
Db	361	GPAPHVLSFLARPMSMDIVSA	FSFLAPLPLSPISVYISVYSGRGKDPDAHVAVNL	419

```
Search completed: March 4, 2003, 12:34:33
Job time : 97.9157 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:31:51 ; Search time 24.3996 Seconds

(without alignments)  
1650.860 Million cell updates/sec

Title: US-09-234-208B-2

Sequence: 1 MEIALCRWGLLLALLPPGA.....VGRGPDPAHVAVNLRYEG 419

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	82.1	1255	1 A24571	protein-tyrosine k
2	1608.5	70.3	1260	1 TVRTNU	protein-tyrosine k
3	1571	68.7	1254	1 I48161	p-185 precursor -
4	796.5	34.8	1223	1 TVCHLV	epidermal growth f
5	794	34.7	527	2 A42032	epidermal growth f
6	793	34.7	1210	1 GCHUE	epidermal growth f
7	789	34.5	1210	2 A53183	epidermal growth f
8	781.5	34.2	644	2 A35325	epidermal growth f
9	775	33.9	1308	2 A47253	epidermal growth f
10	735.5	32.2	1342	2 A36223	kinase-related tra
11	692	30.3	1339	2 JC4387	epidermal growth f
12	681.5	29.8	1166	1 S06142	protein-tyrosine k
13	575.5	25.2	843	2 A27131	epidermal growth f
14	419	18.3	1323	2 E88257	epidermal growth f
15	419	18.3	1374	2 S70712	protein-tyrosine k
16	416	18.2	1369	2 S70713	protein-tyrosine k
17	415	18.1	1330	1 GQFFE	epidermal growth f
18	363.5	15.9	366	2 D45558	epidermal growth f
19	363.5	15.9	1717	1 A45558	epidermal growth f
20	342.5	15.0	1363	2 T43220	insulin-like growt
21	331	14.5	333	2 B45558	epidermal growth f
22	331	14.5	342	2 C45558	epidermal growth f
23	294	12.9	2101	2 S57245	insulin receptor (
24	294	12.9	2148	1 A56081	insulin receptor -
25	291	12.7	1477	2 T18534	protein-tyrosine k
26	278	12.2	1300	2 A35502	insulin receptor-t
27	272	11.9	540	2 B47417	insulin receptor-t
28	269.5	11.8	1607	2 T43212	insulin-like growt
29	263.5	11.5	1382	1 INHUR	insulin receptor p

30	263.5	11.5	1383	2 A36080	insulin receptor p
31	261	11.4	1372	2 A34157	insulin receptor p
32	258	11.3	1390	2 T30346	insulin receptor -
33	257.5	11.3	1367	1 IGHUR1	insulin-like growt
34	252.5	11.0	1371	2 A33837	insulin-like growt
35	249	10.9	1268	2 B36502	insulin receptor-t
36	235	10.3	329	2 A48805	insulin-like growt
37	226.5	9.9	183	2 JH0803	tyrosine kinase re
38	203	8.9	1846	2 T42047	insulin receptor h
39	160.5	7.0	1299	2 T43251	furin (EC 3.4.21.7
40	144	6.3	1548	2 S34583	serine proteinase
41	142	6.2	1959	1 AGRT	agrin - rat
42	136.5	6.0	1111	2 T26972	hypothetical prote
43	133.5	5.8	915	1 A48225	subtilisin-like pr
44	132	5.8	1574	2 T13954	MEGF6 protein - ra
45	131.3	5.7	915	2 B48225	probable proteol

#### ALIGNMENTS

##### RESULT 1

A24571  
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human  
N:Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein e  
C:Species: Homo sapiens (man)  
C:Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999  
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622  
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T  
Nature 319, 230-234, 1986  
A>Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growt  
A:Reference number: A24571; MUID:8618663; PMID:3003577  
A:Accession: A24571  
A:Molecule type: mRNA  
A:Residues: 1-1255 <YAM>  
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985  
A>Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid  
A:Reference number: A25491; MUID:86016729; PMID:2995967  
A:Accession: A25491  
A:Molecule type: DNA  
A:Residues: 737-1031 <SEM>  
A:Cross-references: GB:M11767; NID:9182163; PIDN:AAA35808.1; PID:9553282  
R:Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg  
Science 230, 1132-1139, 1985  
A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro  
A:Reference number: A44188; MUID:86070181; PMID:2999974  
A:Accession: A44188  
A:Molecule type: DNA  
A:Residues: 740-910 <COU1>  
A:Cross-references: GB:M12036; NID:9183988; PIDN:AAA35978.1; PID:9183989  
A:Accession: B44188  
A:Molecule type: mRNA  
A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>  
A:Cross-references: GB:M11730; NID:9183986  
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.  
Science 229, 974-976, 1985  
A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.  
A:Reference number: I59509; MUID:85272597; PMID:2992089  
A:Accession: I59509  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 832-909 <REX>  
A:Cross-references: GB:L29395; NID:9459807; PIDN:AAA35809.1; PID:9459808  
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.  
Mol. Cell. Biol. 7, 2597-2601, 1987  
A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio  
A:Reference number: I57622; MUID:87286898; PMID:3039351  
A:Accession: I57622  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:G183983; PIDN:AAA58637.1; PID:9553332  
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
C:Genetics:  
A:Gene: GDB:ERBB2; NGL: NEU; HER-2  
A:Cross-references: GDB:120613; OMIM:164870  
A:Map position: 17q21.1-17q21.1  
A:Introns: 25/1; 75/3; 147/1; 883/3  
A:Note: the list of introns is incomplete  
C:Function:  
A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
kinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
F:22-653/Domain: extracellular #status predicted <EXT>  
F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>  
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>  
F:654-675/Domain: transmembrane #status predicted <TM>  
F:676-1255/Domain: intracellular #status predicted <INT>  
F:718-983/Domain: protein kinase homology <KIN>  
F:726-734/Region: protein kinase ATP-binding motif  
F:68.124.187.259.530.571.629/Binding site: carboxydrate (Asn) (covalent) #status predict  
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
F:753/Active site: Lys #status predicted  
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 82.1%; Score 1878; DB 1; Length 1255;

Best local Similarity 83.0%; Pred. No. 5.9e-121;

Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```
QY 1 METALACRMGLLALLPRGAASVQVCTGDMKRLRASPETHIDMYRLHYOGCQVVOGNL 60
DB 1 METALACRMGLLALLPRGAASVQVCTGDMKRLRASPETHIDMYRLHYOGCQVVOGNL 60
QY 61 ELTYLPTNLSLFDIOEVOGYVLLAHNOVROVRLQRLRYVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNLSLFDIOEVOGYVLLAHNOVROVRLQRLRYVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVYGASPGSLRELOLSLTLEILKGVLIQNRPOLCYODITIMKDFHKNQOLA 180
DB 121 DPLNNTPTVYGASPGSLRELOLSLTLEILKGVLIQNRPOLCYODITIMKDFHKNQOLA 180
QY 181 LTLIDTNRSRACHPCSPMKSGSRGWESSESDQSLTRTVACAGCARKGRLPTDCCHQD 240
DB 181 LTLIDTNRSRACHPCSPMKSGSRGWESSESDQSLTRTVACAGCARKGRLPTDCCHQD 240
QY 241 AACCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCVTAC 300
DB 241 AACCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCVTAC 300
QY 301 YNYLSTDVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCAR-----GTHSLPRPAAPV 355
DB 301 YNYLSTDVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCGVGLMHNREVRATYSAN 360
QY 356 LRMQPG--PAHPLYSLRLPSMDLVSAFYSLPLAPLSPTVPT-----SPVSGRGPD 405
DB 356 LRMQPG--PAHPLYSLRLPSMDLVSAFYSLPLAPLSPTVPT-----SPVSGRGPD 405
QY 406 --PDAHVAVNLRYEG 419
DB 406 --PDAHVAVNLRYEG 419
QY 418 SLPDLSVFQNLQVIRG 433
DB 418 SLPDLSVFQNLQVIRG 433
```

## RESULT 2

TYRTNU

Protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 11-Jun-1999

C:Accession: A24562; A61204

R:Barjann, C.I.; Hung, M.C.; Weinberg, R.A.

A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A:Reference number: A24562; MUID:86118662; PMID:3945311

A:Accession: A24562

A:Molecule type: mRNA

A:Residues: 1-1260 <BAR>

A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746

R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohe

Cartinogenesis 12, 1975-1978, 1991

A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals n

2-thiazolylformamide or N-methyl-N-nitrosourea.

A:Reference number: A61204; MUID:92035293; PMID:1682063

A:Accession: A61204

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 637-663, 'V', 665-702 <MAS>

A:Note: authors translated the codon GCA for residue 25 as Val

C:Genetics:

A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TM>

F:723-988/Domain: protein kinase homology <KIN>

F:731-739/Region: protein kinase ATP-binding motif

F:71,191,263,535,576,634/Binding site: carboxydrate (Asn) (covalent) #status predicte

F:691/Binding site: phosphate (Thr) (covalent) #status predicted

F:758/Active site: Lys #status predicted

F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 70.3%; Score 1608.5; DB 1; Length 1260;

Best local Similarity 85.0%; Pred. No. 1.8e-102;

Matches 294; Conservative 15; Mismatches 36; Indels 1; Gaps 1;

```
QY 1 METALACRMGLLALLPRGAASVQVCTGDMKRLRASPETHIDMYRLHYOGCQVVOGNL 60
DB 1 METALACRMGLLALLPRGAASVQVCTGDMKRLRASPETHIDMYRLHYOGCQVVOGNL 60
QY 4 MELALACRMGLLALLPRGASVQVCTGDMKRLRASPETHIDMYRLHYOGCQVVOGNL 63
DB 4 MELALACRMGLLALLPRGASVQVCTGDMKRLRASPETHIDMYRLHYOGCQVVOGNL 63
QY 61 ELTYLPTNLSLFDIOEVOGYVLLAHNOVROVRLQRLRYVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNLSLFDIOEVOGYVLLAHNOVROVRLQRLRYVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVYGASPGSLRELOLSLTLEILKGVLIQNRPOLCYODITIMKDFHKNQOLA 179
DB 121 DPLNNTPTVYGASPGSLRELOLSLTLEILKGVLIQNRPOLCYODITIMKDFHKNQOLA 179
QY 181 LTLIDTNRSRACHPCSPMKSGSRGWESSESDQSLTRTVACAGCARKGRLPTDCCHQD 239
DB 181 LTLIDTNRSRACHPCSPMKSGSRGWESSESDQSLTRTVACAGCARKGRLPTDCCHQD 239
QY 241 AACCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCVTAC 299
DB 241 AACCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCVTAC 299
QY 301 YNYLSTDVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCARAGHSL 345
DB 301 YNYLSTDVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCARAGHSL 345
QY 304 PNYVLTSTVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCGYL 349
DB 304 PNYVLTSTVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCARVCGYL 349
```

## RESULT 3

I48161

P-185 precursor - golden hamster

C:Species: Mesocricetus auratus (golden hamster)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999

C:Accession: I48161

R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishika

Gene 140, 251-255, 1994

A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.

A:Reference number: I48161; MUID:94193007; PMID:7908275

A:Accession: I48161

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1254 <RES>

A:Cross-references: GB:D16295; NID:9493236; PIDN:BA03801.1; PID:9747595



Db 139 TO-----GLRELPMKRLSELINGSVKISNNPKLCNMDTVMNDIIDSRSK-PLTVLD 189

OY 186 -TRNSRACHPCSPMKSGSRGSGESSEDCOSLTRVCAGGCA-BCKGPLPTDCHEOCAG 243

Db 190 FASNLSSCPCHCRNCTEDHWCAGAEQNCQTLTVICAQOCSSGRGRCGVPSDDCHNOCAG 249

OY 244 CTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYFGASCVTACPYNY 303

Db 250 CTGPRESDCLACRFRDADATCKDTCPLVLYNPTTYQMDVNPBGKYSFGATVCVCPHNY 309

OY 304 LSTDVSGCTLVCPHNEVTAEDGTORCEKCSKPCAR 340

Db 310 VVTDHSGVRSCTDYEV-EENGVRCKRCKDGLCSK 345

RESULT 6

GOHUE

epidermal growth factor receptor precursor - human

N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Homo sapiens (man)

C:Date: 15-Nov-1984 #sequence\_revision 27-Nov-1985 #text\_change 11-Jun-1999

C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143; A33

R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y

rg, P.H.

Nature 309, 418-425, 1984

A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312

A:Accession: A00641

A:Molecule type: mRNA

A:Residues: 1-1210 <UL>

A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g575924

A:Note: the authors translated the codon AAG for residue 540 as Asn

R:Shih, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.

Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985

A:Title: Characterization and sequence of the promoter region of the human epidermal gro

A:Reference number: A25772; MUID:85270438; PMID:2591899

A:Accession: A25772

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-29 <ISH>

A:Cross-references: GB:ML1234; NID:g181981; PIDN:AA53370.1; PID:g553272

R: Haley, J.; Whitfield, N.; Bennett, P.; Kitchington, D.; Ulrich, A.; Waterfield, M.

Oncogene Res. 1, 375-396, 1987

A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification c

A:Reference number: S30024; MUID:88217333; PMID:3329716

A:Accession: S30024

A:Molecule type: DNA

A:Residues: 1-29 <HA>

A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119

R: Haley, J.D.; Waterfield, M.D.

J. Biol. Chem. 266, 1746-1753, 1991

A:Title: Contributory effects of de Novo transcription and premature transcript terminat

A:Reference number: A38672; MUID:91107677; PMID:1988448

A:Accession: A38672

A:Molecule type: DNA

A:Residues: 1-29 <HAL>

A:Cross-references: GB:M38425; NID:g181977; PIDN:AA63171.1; PID:g553271

A:Experimental source: carcinoma cell line A431-7

R: Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Mer

Nature 309, 806-810, 1984

A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs

A:Reference number: A00642; MUID:84245835; PMID:6330563

A:Accession: A00642

A:Molecule type: mRNA

A:Residues: 'RCAMRR', 150-187, 'KSVIQA', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321

., 798-799, 'TD', 802-811, 'R', 813-942 <UY>

A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF recep

R: Lin, C.R.; Chen, W.S.; Krutiger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.

Science 224, 843-848, 1984

A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification

A:Reference number: A43615; MUID:84196372; PMID:6526261

A:Accession: A43615

A:Molecule type: mRNA

A:Residues: 713-964 <LIN>

A:Experimental source: epidermoid carcinoma cell line A431

R: Simmen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.

Biochem. Biophys. Res. Commun. 124, 125-132, 1984

A:Reference number: A23062; MUID:85046483; PMID:6093780

A:Accession: A23062

A:Molecule type: mRNA

A:Residues: 1028-1210 <SIM>

R: Weber, W.; Gill, G.N.; Speiss, J.

Science 224, 294-297, 1984

A:Reference number: A05281; MUID:84172183; PMID:6324343

A:Accession: A05281

A:Molecule type: protein

A:Residues: 25-30, 'S', 32-51; 454-467 <WEB>

R: Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.

J. Biol. Chem. 260, 5205-5208, 1985

A:Title: Identification of residues in the nucleotide binding site of the epidermal g

A:Reference number: A60143; MUID:85182650; PMID:2985580

A:Accession: A60143

A:Molecule type: protein

A:Residues: 740-744, 'X', 746-747 <RUS>

R: Mroczkowski, B.; Mosig, G.; Cohen, S.

Nature 309, 270-273, 1984

A:Title: ATP-stimulated interaction between epidermal growth factor receptor and supe

A:Reference number: A38023; MUID:84191554; PMID:6325948

A:Contents: annotation; receptor activity

A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA

R: Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C

Cell 59, 33-43, 1989

A:Title: Functional independence of the epidermal growth factor receptor from a doma

A:Reference number: A33331; MUID:90003233; PMID:2790960

A:Contents: annotation; internalization signal

C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-recepto

C:Genetics:

A:Gene: GDB:EGFR

A:Cross-references: GDB:120610; OMIM:131550

A:Map position: 7p12.3-7p12.1

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-1210/Product: EGF receptor #status predicted <MAT>

F:25-645/Domain: extracellular #status predicted <EXT>

F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>

F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>

F:646-668/Domain: transmembrane #status predicted <TM>

F:669-1210/Domain: intracellular #status predicted <INT>

F:710-975/Domain: protein kinase homology <KIN>

F:718-726/Region: protein kinase ATP-binding motif

F:999-1046/Region: coated-pit mediated internalization signal

F:1047-1210/Region: inhibitory

F:128, 175, 352, 413, 444, 528, 603/Binding site: carbohydrate (asn) (covalent) #status pre

F:745/Active site: Lys #status experimental

Query Match 34.7%; Score 793; DB 1; Length 1210;

Best Local Similarity 45.3%; Pred. No. 1,3e-46;

Matches 151; Conservative .50; Mismatches 118; Indels 14; Gaps 4;

OY 11 LLLALLPRGAA--STGYCTGDMKLLRASPETHLDMLRLHYGGCGVQGNLELTLPIN 68

Db 14 LLAALDPASRALEERKVCOTSNKRLQLGTGFEDHPLSLDMFNFNCEVAGNLEITVQNR 73

OY 69 ASLSFDLIDIEVGYVLIHNOVROYPLRLRIVRGTOLEEDNVYALVDNGDPLNNTTP 128

Db 74 YDLSFLKTDIEVAGVYLIHNTYERIPLEQLQIRGNMYENSYALAVLSND----- 126

OY 129 VTGASFGGLEQLRSLTETLKGVLIQRNPOLCYODTILMKDIFRKNQOLATLIDTNR 188

Db 127 ---ANKTGKLELPMRNLOETILHGAVFSSNPALCNVESIQWRDVSDFLSSMSFOQNH 183

OY 189 SRACHCSPMKSGSRGSGESSEDCOSLTRVCAGGCA-RKGPLPTDCHEOCAGACTGP 247

Db 184 LGSQCRKDCPSPNGSCWAGAEQNCQTLTVICAQOCSSGRGRCGVPSDDCHNOCAGCTGP 243



Db 236 CAAGCTGPRESDCLVCHRRFDEATCKDTCPPMLTYNPPTYQMDVNEGKYSFGATCVKRC 295  
 QY 300 PNYLSTDVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCAR 340  
 Db 296 PNNYVYTDHSGCVACGCPDYEV-EDGVSCKCKCGPCRK 335

## RESULT 9

A47253  
 A:Gene: epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)  
 C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999

C:Accession: A47253  
 R:Plowman, G.D.; Clouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; NE

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993  
 A:Title: Ligand-specific activation of HERK/p180erbB4, a fourth member of the epidermal

A:Reference number: A47253; MUID:93189574; PMID:9383326  
 A:Accession: A47253

A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid

A:Residues: 1-1308 <PIG>  
 A:Cross-references: GB:I07868; NID:9337359; PIDN:AA59446.1; PID:9337360

A:Note: sequence extracted from NCBI backbone (NCBIP:126842)  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; growth factor receptor  
 F:716-981/Domain: protein kinase homology <KIN>  
 F:724-732/Region: protein kinase ATP-binding motif

Query Match 33.9%; Score 775; DB 2; Length 1308;  
 Best Local Similarity 45.1%; Pred. No. 2.4e-45;  
 Matches 157; Conservative 45; Mismatches 124; Indels 22; Gaps 7;

QY 9 WGLLALLPPGAA---STOVCTGTDMLRLPASPEHMDLRHLVGGGOVQGNLELY 64  
 Db 8 WWSVLLVAGTYQPSDSOSVCGATENKLSLSLEQYRALRKYNEVWGNLEITS 67  
 QY 65 LPTNASLFLQDIOEVQGYVLAHNOVROVPLQRLRIYGTQLFEDNYALAVLDNGDPLN 124  
 Db 68 IEHNRDLFSRVYEVYVLAHNOVRYLPLENRLIRGTQLYEDRYALAIPLNRYKDG 127  
 QY 125 NTPPYVGASPGGLRELQSLSTFLILKGVLIQGNPOLCYODTILMKDIFHKNNQALTL 184  
 Db 128 NF-----GLQELGLNLTLEILNGVYVDDNKFCLYADTTHMODIVRNWPSMLTLV 178  
 QY 185 DPNRSRACHPCSPMGKSGRSGESSEDCOSLRTVCAGGC-ARCKGPLPTDCHCEGCAAG 243  
 Db 179 STNGSSGCGRCHSGCTG-KCWGPTENHCQTLRTVCACBQCDGRCYPRYVSDCHRECAAG 237  
 QY 244 CTGPKHSDCLAHFNHSGICELHCPALVYNTDFESMPNPEGRYTFGASCVTACPYNY 303  
 Db 238 CSGPKDTDFACNMFNDSCACVQCQPTFVYNPFTQLFENFNKATYTGAFVCKPCPNHF 297  
 QY 304 LSTDVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCAR-----GTHSL 346  
 Db 298 V-VDSSSCVACPPSSKMEV-EENGIMKRCPCDIDCPKACDGIQTGLSM 343

## RESULT 10

A36223  
 A:Gene: kinase-related transforming protein (erbB3) (BC 2.7.1.-) precursor - human

C:Species: Homo sapiens (man)  
 C:Date: 04-Oct-1991 #sequence\_revision 13-Jan-1993 #text\_change 17-Nov-2000

C:Accession: A36223; MUID:90083234; PMID:2687875  
 R:Kraus, M.H.; Issing, W.; Mikl, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989  
 A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g

A:Reference number: A36223  
 A:Accession: A36223

A:Status: preliminary  
 A:Molecule type: mRNA

A:Residues: 1-1342 <KRA>  
 A:Cross-references: GB:M29366  
 R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J.

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990  
 A:Title: Molecular cloning and expression of another epidermal growth factor receptor

A:Reference number: 159164; MUID:90311312; PMID:2164210  
 A:Accession: 159164

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA

A:Residues: 1-559, 'G', '561-957', 'F', '959-1063', 'G', '1065-1342 <RES>  
 A:Cross-references: GB:M34309; NID:9183990; PIDN:AAA35979.1; PID:9306841

C:Genetics:  
 A:Gene: GDB:ERBB3; HER3

A:Cross-references: GDB:119880; OMTM:190151  
 A:Map position: 12q13-12q13

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom  
 C:Keywords: ATP; phosphotransferase

F:707-972/Domain: protein kinase homology <KIN>  
 F:715-723/Region: protein kinase ATP-binding motif

Query Match 32.2%; Score 735.5; DB 2; Length 1342;  
 Best Local Similarity 44.0%; Pred. No. 1.3e-42;  
 Matches 150; Conservative 49; Mismatches 121; Indels 21; Gaps 9;

QY 10 GLLALLPPGAA--STOVCTGTDMLRLPASPEHMDLRHLVGGGOVQGNLELYPT 67  
 Db 11 GLFSLARGSEVGNQAVCPGLNGLSVYGDENQYQTLKYERCEVWGNLEIYLGH 70  
 QY 68 NLSLFLQDIOEVQGYVLAHNOVROVPLQRLRIYGTQLFEDNYALAVLDNGDPLNNT 127  
 Db 71 NADLSFLQIREVYGVLAHNEFSTPLPNRYRGTOYVYGRKAIPTM-----LNNT 125  
 QY 128 PYTGASPGGLRELQSLSTFLILKGVLIQGNPOLCYODTILMKDIFHKNNQALTL 187  
 Db 126 ---NSSHALROLRLTLQLEILSGVYIEKNDKLCMDTIDWRDIYRDD--AEIYVD 178  
 QY 188 RSRACHPCSPMGKSGRSGESSEDCOSLRTVCAGGC-ARCKGPLPTDCHCEGCAAGCTG 246  
 Db 179 NGRSCPPCHEVCKG-KCWGPGSEDCQTLTKTCAPDCNCHGCPNMQCHCEGAGCGS 237  
 QY 247 PKHSDCLAHFNHSGICELHCPALVYNTDFESMPNPEGRYTFGASCVTACPYNYLST 306  
 Db 238 PQDTDFACRHNHNDSCACVPCRPOLYVYKLFQLEPNHNTYQYGVGVACPNFV-V 296  
 QY 307 DVSGCTLVCPRLHNOEYTAEDGTORCEKSKPCAR---GTHS 344  
 Db 297 DQTSVCVACPPDKMEVD--KNGLMCEPCGCLCPKACEGTGS 336

## RESULT 11

JC4387  
 A:Gene: epidermal growth factor receptor homolog precursor - rat

N:Alternate names: ErbB3 protein; HER3 protein  
 C:Species: Rattus norvegicus (Norway rat)

C:Date: 17-Jan-1996 #sequence\_revision 19-Apr-1996 #text\_change 13-Nov-1998  
 C:Accession: JC4387

R:Hellier, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.  
 Gene 165, 279-284, 1995

A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protei  
 A:Reference number: JC4387; MUID:96096535; PMID:8522190

A:Accession: JC4387  
 A:Molecule type: mRNA

A:Residues: 1-1339 <HEL>  
 A:Cross-references: GB:U29339; NID:9915389; PID:9915390

A:Experimental source: liver  
 A:Note: The authors translated the codon AAG for residue 369 as Thr and GGT for resid

C:Comment: This protein is a functional heregulin receptor that transduces signals to  
 C:Genetics:

A:Gene: ErbB3  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein  
 F:20-1339/Product: epidermal growth factor homolog <status predicted <MAY>

F:640-659/Domain: transmembrane #status predicted <MAY>  
 F:705-970/Domain: protein kinase homology <KIN>  
 F:713-721/Region: protein kinase ATP-binding motif







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:25:15 ; Search time 15.1446 seconds  
(without alignments)  
1147.512 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287  
Sequence: 1 MEIALCRMGLLALLPFGA.....VGRGPDPAHVAVNLSRYEG 419

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	82.1	1255	1 ERB2_HUMAN	P04626 homo sapien
2	1608.5	70.3	1257	1 ERB2_RAT	P06494 rattus norv
3	1571	68.7	1254	1 ERB2_MESAU	O60553 mesocricetu
4	796.5	34.8	703	1 EGFR_CHICK	P13387 gallus gall
5	793	34.7	1210	1 EGFR_HUMAN	P00533 homo sapien
6	789	34.5	1210	1 EGFR_MOUSE	O01279 mus musculu
7	779.5	34.1	1308	1 ERB4_RAT	O62956 rattus norv
8	775	33.9	1308	1 ERB4_HUMAN	O15303 homo sapien
9	735.5	32.2	1342	1 ERB3_HUMAN	P21860 homo sapien
10	698	30.5	1339	1 ERB3_RAT	O62799 rattus norv
11	681.5	29.8	1167	1 XMRK_XIPMA	P13388 xiphophorus
12	574.5	25.1	1426	1 EGFR_DROME	P04412 drosophila
13	419	18.3	1323	1 LT23_CABEL	P24348 caenorhabdi
14	342.5	15.0	1363	1 ILPR_BRALA	O02466 branchiosto
15	294	12.9	2146	1 INSR_DROME	P09208 drosophila
16	291	12.7	1477	1 HTK7_HYDAT	O25197 hydra atten
17	290	12.7	1300	1 IRR_MOUSE	O9wt14 mus musculu
18	278	12.2	1300	1 IRR_CAVPO	P14617 cavia porce
19	270	11.8	581	1 IRR_RAT	O64716 rattus norv
20	269.5	11.8	1607	1 MIPR_LYMTST	O25410 lymphaea sta
21	269	11.8	1297	1 INSR_HUMAN	P14616 homo sapien
22	264.5	11.6	1382	1 INSR_HUMAN	P06213 homo sapien
23	263.5	11.5	1383	1 INSR_RAT	P15127 rattus norv
24	261	11.4	1372	1 INSR_MOUSE	P15208 mus musculu
25	258	11.3	1390	1 INSR_AEDAE	O93105 aedes aegypt
26	257.5	11.3	1367	1 IGR_HUMAN	P08069 homo sapien
27	252.5	11.0	1370	1 IGR_RAT	P24062 rattus norv
28	250.5	11.0	1373	1 IGR_MOUSE	O60701 mus musculu
29	169	7.4	1696	1 PKC5_BRACL	O9n113 branchiosto
30	144	6.3	1877	1 PKC5_MOUSE	O04552 mus musculu
31	142	6.2	1959	1 AGRI_RAT	P25304 rattus norv
32	131.5	5.7	913	1 PKC5_HUMAN	O92824 homo sapien
33	129.5	5.7	1680	1 FUR2_DROME	P30432 drosophila

34	128	5.6	830	1 SREC_HUMAN	O14162 homo sapien
35	124.5	5.4	1877	1 PKC5_RAT	P41413 rattus norv
36	123.5	5.4	937	1 PAC4_RAT	O63415 rattus norv
37	123	5.4	417	1 TR12_HUMAN	O93038 h tumor nec
38	120	5.2	498	1 TNFR_MOUSE	O60846 mus musculu
39	120	5.2	3695	1 LMAS_HUMAN	O15230 homo sapien
40	119.5	5.2	667	1 TS11_GIALA	O03185 giardia lam
41	119	5.2	2282	1 ZAN_RABIT	P57999 oryctolagus
42	118.5	5.2	551	1 LEM2_RABIT	P27113 oryctolagus
43	118.5	5.2	3718	1 LMAS_MOUSE	O61001 mus musculu
44	117	5.1	469	1 PROP_HUMAN	P27918 homo sapien
45	117	5.1	3075	1 LMAL_HUMAN	P25391 homo sapien

## ALIGNMENTS

RESULT 1  
ERB2\_HUMAN STANDARD: PRT; 1255 AA.  
ID ERB2\_HUMAN  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)  
DE (P185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell  
DE surface receptor HER2) (M19.19).  
GN ERB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=6118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.;  
RT "Similarity of protein encoded by the human c-erb-B-2 gene to  
RT epidermal growth factor receptor.";  
RL Nature 319:230-234(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=66070181; PubMed=2999974;  
RA Cussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Francke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RN Science 230:1132-1139(1985).  
RN [3]  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE=86016729; PubMed=2995967;  
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;  
RT A v-erbB-related protooncogene, c-erbB-2, is distinct from the  
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a  
RT human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
RN [4]  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429(1993).  
RN [5]  
RP FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
RP ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
RP POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
RP ALPHA AND AMPHIREGULIN.  
CC CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein  
CC tyrosine phosphatase.  
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
CC RESIDUES (BY SIMILARITY).  
CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
CC POSITIONS 654 AND 655. ALLELE B1 (654-TLE-TLE-655) HAS A FREQUENCY  
CC OF 0.782; ALLELE B2 (654-TLE-VAL-655) HAS A FREQUENCY OF 0.206;  
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M11767; AAA35808.1; -  
DR EMBL: M11761; AAA35808.1; JOINED.  
DR EMBL: M11762; AAA35808.1; JOINED.  
DR EMBL: M11763; AAA35808.1; JOINED.  
DR EMBL: M11764; AAA35808.1; JOINED.  
DR EMBL: M11765; AAA35808.1; JOINED.  
DR EMBL: M11766; AAA35808.1; JOINED.  
DR EMBL: M11730; AAA75493.1; -  
DR EMBL: M12036; AAA35978.1; -  
DR EMBL: X03363; CAA27060.1; -  
DR PIR: A25491; A25491.  
DR PIR: A24571; A24571.  
DR HSSP: P11362; 1FGK.  
DR Genew: HGNC:3430; ERBB2.  
DR MIM: 164870; -  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_kinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_kinase.  
DR InterPro: IPR004019; YLP\_motif.  
DR Pfam: PF00069; kinase.1.  
DR Pfam: PF00757; Furin-like.1.  
DR Pfam: PF01030; Recep\_L\_domain.2.  
DR Pfam: PF02757; YLP.2.  
DR ProDom: PD000001; Euk\_kinase.1.  
DR SMART: SM00261; FU.3.  
DR SMART: SM00219; TYRC.1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP.1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR.1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM.1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW Polymorphism.  
KW -----  
FT CHAIN 1 21 POTENTIAL.  
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 653 675 POTENTIAL.  
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 720 987 PROTEIN KINASE.  
FT NE\_BIND 726 734 ATP (BY SIMILARITY).  
FT BINDING 753 753 ATP (BY SIMILARITY).  
FT ACT\_SITE 845 845 BY SIMILARITY.  
FT DISULFID 195 204 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 220 227 BY SIMILARITY.  
FT DISULFID 224 235 BY SIMILARITY.  
FT DISULFID 236 244 BY SIMILARITY.  
FT DISULFID 240 252 BY SIMILARITY.  
FT DISULFID 255 264 BY SIMILARITY.  
FT DISULFID 268 295 BY SIMILARITY.  
FT DISULFID 299 311 BY SIMILARITY.  
FT DISULFID 315 331 BY SIMILARITY.  
FT DISULFID 334 338 BY SIMILARITY.  
FT DISULFID 511 520 BY SIMILARITY.  
FT DISULFID 515 528 BY SIMILARITY.  
FT DISULFID 531 540 BY SIMILARITY.  
FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
FT DISULFID 567 584 BY SIMILARITY.  
FT DISULFID 587 596 BY SIMILARITY.  
FT DISULFID 600 623 BY SIMILARITY.  
FT DISULFID 626 634 BY SIMILARITY.  
FT DISULFID 630 642 BY SIMILARITY.  
FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).  
FT VARIANT 654 654 I -> V.  
FT VARIANT 654 654 /FTID-VAR\_004077.  
FT VARIANT 655 655 I -> V.  
FT VARIANT 655 655 /FTID-VAR\_004078.  
FT CONFLICT 1170 1170 P -> A (IN REF. 2).  
SQ SEQUENCE 1255 AA: 137909 MW: 39899DDA04DC962 CRC64;  
Query Match 82.1%; Score 1878; DB 1; Length 1255;  
Best Local Similarity 83.0%; Pred. No. 6,7e-138;  
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;  
QY 1 METALCRWGLLALPPGASSTQVCTGDMKRLRPSPEHIDMLRHLVGGCGVGNL 60  
Db 1 METALCRWGLLALPPGASSTQVCTGDMKRLRPSPEHIDMLRHLVGGCGVGNL 60  
QY 61 ELTYLPTNASISFLQDIQEVGYVLIANOVROYPLQRLRIVRGTOLEEDNYALAVDNG 120  
Db 61 ELTYLPTNASISFLQDIQEVGYVLIANOVROYPLQRLRIVRGTOLEEDNYALAVDNG 120  
QY 121 DPLNNTPTVVGASFGGRLREILKGVLIQNPOLCQDTITLMKDFHKNQOLA 180  
Db 121 DPLNNTPTVVGASFGGRLREILKGVLIQNPOLCQDTITLMKDFHKNQOLA 180  
QY 181 LTLIDTKRSRACHPCSPMGSGRMCWGSSEDCOSLFTVAGGARGKGLPDDCCEQC 240  
Db 181 LTLIDTKRSRACHPCSPMGSGRMCWGSSEDCOSLFTVAGGARGKGLPDDCCEQC 240  
QY 241 AAGCTGPKHSDCIACLFHNSGICELHCPALVYNTDTPESMPNDEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCIACLFHNSGICELHCPALVYNTDTPESMPNDEGRYTFGASCVTACP 300  
QY 301 YNYISTDVSGCTLVCPHLNDEVTAEDCTQRCCKSKPCAR----GHSLLPPRAAVPVP 355  
Db 301 YNYISTDVSGCTLVCPHLNDEVTAEDCTQRCCKSKPCARCYGLMEHLREVARAVTSAN 360  
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLLPLAPLSTSVPI-----SPVSYGSGPD 405  
Db 361 IQEPAGCKKTFGSLAFLEPSFDGPPASNT---APLOPEOLVFEETLEITGYLISAMPD 417  
QY 406 --PDAAVAVNLRYEG 419  
Db 418 SLDPDLSVFQMLQYIRG 433  
RESULT 2  
ERBB2\_RAT  
ID ERBB2\_RAT STANDARD; PRT; 1257 AA.  
AC P06484;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor  
DE receptor-related protein).  
OS ERBB2 OR NEU.  
GN Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Neuroblastoma;  
RX MEDLINE=86118662; PubMed=3945311;  
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;  
RT "The neu oncogene encodes an epidermal growth factor receptor-related protein."  
RL Nature 319:226-230(1986).  
RN [2]  
RP SEQUENCE OF 852-905 FROM N.A.  
RC TISSUE-Sciatic nerve;  
RX MEDLINE=91222560; PubMed=2025425;  
RA Lai C., Lemke G.;  
RT "An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system."  
RL Neuron 6:691-704(1991).  
RN [3]  
RP STRUCTURE BY NMR OF 650-668.  
RX MEDLINE=92155181; PubMed=1346763;  
RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D., Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;  
RT "Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein."  
RL EMBO J. 11:43-48(1992).  
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMHIREGULIN.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.  
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS. THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X03362; CAA27059.1; ALT\_INIT.  
DR PIR: A24562; TVRTU.  
DR HSSP: P11362; IFGK.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR InterPro: IPR004019; YLP\_molif.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR Pfam: PF02757; YLP; 2.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00261; Euk\_3.  
DR SMART: SM00219; Tyrc; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
KM Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;  
KM Transferase: Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KM Proto-oncogene; Disease mutation.  
FT SIGNAL 1 21  
FT CHAIN 22 1257  
FT DOMAIN 22 654  
FT TRANSMEM 655 677  
FT DOMAIN 678 1257  
FT DOMAIN 159 369  
CYTOPLASMIC (POTENTIAL).  
CYS-RICH.

FT	DOMAIN	473	646	CYS-RICH.
FT	DOMAIN	722	989	PROTEIN KINASE.
FT	NP_BIND	728	736	ATP (BY SIMILARITY).
FT	BINDING	755	755	ATP (BY SIMILARITY).
FT	ACT_SITE	847	847	BY SIMILARITY.
FT	DISULFID	196	205	BY SIMILARITY.
FT	DISULFID	200	213	BY SIMILARITY.
FT	DISULFID	221	228	BY SIMILARITY.
FT	DISULFID	225	236	BY SIMILARITY.
FT	DISULFID	237	245	BY SIMILARITY.
FT	DISULFID	241	253	BY SIMILARITY.
FT	DISULFID	256	265	BY SIMILARITY.
FT	DISULFID	269	296	BY SIMILARITY.
FT	DISULFID	300	312	BY SIMILARITY.
FT	DISULFID	316	332	BY SIMILARITY.
FT	DISULFID	335	339	BY SIMILARITY.
FT	DISULFID	513	522	BY SIMILARITY.
FT	DISULFID	517	530	BY SIMILARITY.
FT	DISULFID	533	542	BY SIMILARITY.
FT	DISULFID	546	562	BY SIMILARITY.
FT	DISULFID	565	578	BY SIMILARITY.
FT	DISULFID	569	586	BY SIMILARITY.
FT	DISULFID	589	598	BY SIMILARITY.
FT	DISULFID	602	625	BY SIMILARITY.
FT	DISULFID	628	636	BY SIMILARITY.
FT	DISULFID	632	644	BY SIMILARITY.
FT	MOD_RES	1141	1141	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1250	1250	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	188	188	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	260	260	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	573	573	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	631	631	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	661	661	V -> E (IN ONCOGENIC NEU).
SQ	SEQUENCE	1257 AA;	138831 MM;	6129264583011402 CRC64;

Query Match 70.3%; Score 1608.5; DB 1; Length 1257;  
Best Local Similarity 85.0%; Pred. No. 5.3e-117;  
Matches 294; Conservative 15; Mismatches 36; Indels 1; Gaps 1;

QY	1	MELALACRMGLLALLPRGASTVCTGTDMLRASPETHLDMRLHLYOGCQVVGML	60
DB	1	MELALACRMGLLALLPRGASTVCTGTDMLRASPETHLDMRLHLYOGCQVVGML	60
QY	61	ELTYLPNASTSLFDIOEYGVLLAHNOVROVPRQRLRYVAGTQLFEDNYALAVLDMG	120
DB	61	ELTYLPNASTSLFDIOEYGVLLAHNOVROVPRQRLRYVAGTQLFEDNYALAVLDMR	120
QY	121	DPLNNTTPVT-GASPGRLRELQRLSTELLKGVLLAQRNPOLCYODTILMKDIFHKNNOL	179
DB	121	DPDQNVAASTPGRTPGRLQRLSTELLKGVLLAQRNPOLCYODTILMKDIFHKNNOL	180
QY	180	ALTLIDTNRSRACHPSCPMCKSGCWGESSEDCQSLTRVACAGCARCKGPLPTDCCHQ	239
DB	181	APVDIDTNRSRACHPACACKDKWHGESPDCQILGTFTGSCARGRLPTDCCHQ	240
QY	240	CAAGCTGPRKHSDDLACLFHNSGICELHCPALVTYMTDFEESMHNDEGRFTGASCVC	300
DB	241	CAAGCTGPRKHSDDLACLFHNSGICELHCPALVTYMTDFEESMHNDEGRFTGASCVC	300
QY	300	PYVYLTGDSCTLVCPRLHNOEYTAEDGTPORCEKSKPCARVHSL	345
DB	301	PYVYLTGDSCTLVCPRLHNOEYTAEDGTPORCEKSKPCARVYGL	346

RESULT 3  
ID ERB2\_MESAU STANDARD; PRT: 1254 AA.  
AC 060553;  
DT 15-DEC-1998 (rel. 37, Created)  
DT 15-DEC-1998 (rel. 37, Last annotation update)  
DT 15-JUN-2002 (rel. 41, Last annotation update)

DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 DE (p18erbB2) (NEU proto-oncogene) (C-erbB-2).  
 GN ERBB2 OR NEU.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 NC NCBL\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Nerve;  
 RA MEDLINE=94193007; PubMed=7908275;  
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,  
 RA Yamazaki Y., Ishikawa T.;  
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";  
 RL Gene 140:251-255(1994).  
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine -> ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: D16295; BAA03801.1; -  
 CC DR HSSP: P11362; 1FGK.  
 DR InterPro: IPR000494; EGFR\_L.domain.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recept\_L.domain; 2.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00261; FU; 3.  
 DR SMART: SM00219; TYKc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Proto-oncogene; Disease mutation.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 1 21 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 1254 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 653 675 POTENTIAL.  
 FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 158 368 CYS-RICH.  
 FT DOMAIN 472 644 CYS-RICH.  
 FT DOMAIN 720 987 PROTEIN KINASE.  
 FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
 FT BINDING 753 753 ATP (BY SIMILARITY).  
 FT ACT\_SITE 845 845 BY SIMILARITY.  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.

FT DISULFID 268 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.  
 FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 BY SIMILARITY.  
 FT MOD\_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).  
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).  
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BB1 CRC64;  
 Query Match 68.7%; Score 1571; DB 1; Length 1254;  
 Best Local Similarity 74.4%; Pred. No. 4,2e-114;  
 Matches 299; Conservative 26; Mismatches 67; Indels 10; Gaps 3;  
 QY 1 MEALALCRWGLLLALPPGAATVCTGDMKRLRASPETHIDMLRHLXOGGVQGNL 60  
 DB 1 MELANCGWLLALLSPGASGVCTGDMKRLRASPETHIDYRHLXOGGVQGNL 60  
 QY 61 ELTYLPNMSLSFLQDIQEVQGVYLAHNOVROVPLQRLRIVRGTOLEFDNVALAVLNG 120  
 DB 61 ELTYLPANATLSFLQDIQEVQGYMLIAHSQVRHVPRLQRLRIVRGTOLEFDKVALAVLDNR 120  
 QY 121 DPLNNTPTVVGASPGGIREQLRLSTREILKGVLIORNPOLCYODPIIMKDFHKNNOA 180  
 DB 121 DPLDNVTTATGRTPEGRLEQLRLSTREILKGVLIORNPOLCYODTVLMDVFRKNNOA 180  
 QY 181 LTLIDTNRSRACHPSCSPMKSGRWCSESSDCLSTRVACAGSCACAKGKPLPDCCHQC 240  
 DB 181 PVDIDNRSRACPPACPACKDNHMCWASPEDCGTLGTIAPRAVPAARARLPDCCHEQC 240  
 QY 241 AAGCTGPKHSDCIACLFHNHSGICEIHCPLVYNTDFESMPNPGRYTFGASCVYACP 300  
 DB 241 AAGCTGPKHSDCIACLFHNHSGICEIHCPLVYNTDFESMPNPGRYTFGASCVYTCP 300  
 QY 301 YNLTSTDVGSCITVCPPLHNOEVAEDGTORCEKSCRPACR-----GTHSLPRAVAVP 355  
 DB 301 YNLTSTVGSCITVCPPLNNOEVAEDGTORCEKSCSASCACVCGIAGMHLRGARALITSAN 360  
 QY 356 LRMPG--PAHPVLSFLRPSMDLSVAFSLPLAPLSPTSVPI 395  
 DB 361 IQEPAGCKKIFGSLAFLEPESFD---GNPSSGIALPLPEQIQV 399  
 RESULT 4  
 ID EGFR\_CHICK STANDARD; PRT: 703 AA.  
 AC P13387;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)  
 DE (Fragment).  
 GN EGFR  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;



RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE-Placenta;  
 RX MEDLINE=97078686; PubMed=8918811;  
 RA Reiter J.L., Maithe N.J.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 factor receptor gene encodes a truncated form of the receptor.";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN [14]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE-Placenta;  
 RX MEDLINE=97256547; PubMed=9103388;  
 RA Tleks J.V., Garfili J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 protein (TEGFR) in ovarian cancer.";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN [15]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RC TISSUE-Placenta;  
 RX MEDLINE=21100872; PubMed=11161793;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Maithe N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 mouse alternative Egfr transcripts encoding truncated receptor  
 isoforms.";  
 RL Genomics 71:1-20(2001).  
 RN [16]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schehl C.M.,  
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,  
 RA Maithe N.J.;  
 RT "Human and mouse alternative Egfr transcripts encoding only the  
 extracellular domain of the receptor.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [17]  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krutiger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RT "Expression cloning of human EGF receptor complementary DNA: gene  
 amplification and three related messenger RNA products in A431  
 cells.";  
 RL Science 224:843-848(1984).  
 RN [18]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 variety of RNAs overproduced in A431 carcinoma cells.";  
 RL Nature 309:806-810(1984).  
 RN [19]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046463; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 receptor cDNA from human A431 carcinoma cells.";  
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).  
 RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whitte N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 identification of sequences regulating its transcription.";  
 RL Oncogene Res. 1:375-396(1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature

RT transcript termination in the regulation of human epidermal growth  
 factor receptor proto-oncogene RNA synthesis.";  
 RL J. Biol. Chem. 266:1746-1753(1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 epidermal growth factor receptor gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN [14]  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor  
 and supercoiled DNA.";  
 RL Nature 309:270-273(1984).  
 RN [15]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Iax I., Kris R., Dombalagian M., Honegger A.M.,  
 RA Howk R., Glyvol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 receptor and HER2/neu are located in their carboxyl-terminal tails.  
 Identification of a novel site in EGF receptor.";  
 RL J. Biol. Chem. 264:10667-10671(1989).  
 RN [16]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96398132; PubMed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renout D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 the epidermal growth factor receptor expressed in Chinese hamster  
 ovary fibroblasts.";  
 RL Growth Factors 13:121-132(1996).  
 RN [17]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731668;  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-Oligosaccharides attached to the atypical  
 Asn-X-Cys sequence of recombinant human epidermal growth factor  
 receptor.";  
 RL J. Biochem. 127:65-72(2000).  
 RN [18]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; PubMed=9556602;  
 RA Abe Y., Oda M., Inagaki F., Iax I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor.";  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RN [19]  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 mitogens.";  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is  
 involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine -> ADP + protein  
 tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 alternative splicing.





```

DR PROSITE: P500107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: P500109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: P550011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane. Glycoprotein. Receptor. Signal. Transferase.
KW Tyrosine-protein kinase. ATP-binding. Phosphorylation. Repeat.
FT SIGNAL 1 24
FT CHAIN 25 1210
FT DOMAIN 25 647 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT TRAMEMEM 648 670 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 671 1210 POTENTIAL.
FT REPEAT 75 300 CYTOPLASMIC (POTENTIAL).
FT REPEAT 390 600 APPROXIMATE.
FT DOMAIN 1028 1071 APPROXIMATE.
FT DOMAIN 714 981 SER-RICH.
FT NP_BIND 720 728 PROTEIN KINASE.
FT BINDING 747 747 ATP (BY SIMILARITY).
FT ACT_SITE 839 839 BY SIMILARITY.
FT DISULFID 190 199 BY SIMILARITY.
FT DISULFID 194 207 BY SIMILARITY.
FT DISULFID 215 223 BY SIMILARITY.
FT DISULFID 232 231 BY SIMILARITY.
FT DISULFID 232 240 BY SIMILARITY.
FT DISULFID 226 248 BY SIMILARITY.
FT DISULFID 251 260 BY SIMILARITY.
FT DISULFID 264 291 BY SIMILARITY.
FT DISULFID 295 307 BY SIMILARITY.
FT DISULFID 311 336 BY SIMILARITY.
FT DISULFID 329 333 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 530 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 628 617 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20M6D22D2P5 CRC64;

```

```

Query Match 34.5%; Score 789; DB 1; Length 1210;
Best Local Similarity 46.3%; Pred. No. 1.7e-53;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;

```

```

QY 11 LLLALLPRAA--STOVCTGDMKRLPASPEHLDMLRLHLYOGCOVYOGNLELYLPIN 68
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 14 LLTALCAAGAAEEKVCVCGTSNRLTQLCTFEDHFLSLQRMYNNECVGLNETVYQBN 73
QY 69 ASLSFDIOEVOGYVLIHNOVOVPLQRLRIVRGTOLEFEDVYALAVDNGDPLNTPR 128
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 74 YLSEFLKTQEVAGVYLIHLNVTVERFLENLOIRGNALYENTYALAILSN----- 124
QY 129 VTGASPGGLRELQLSLTELKGLVLIQRNPOLCYODTILMKDI-----PHKNQIALTLI 184

```

```

DB 125 -GNTNRTGLRELPMRLQETILIGAVFSNNPILICNMNDTQWRDIQVWNMSMDL--- 180
QY 185 DNNRSACHPGCPKCGKSGKMGSESDCOSLTRYCAGCA-RCKGPLPTDCHQCAAG 243
DB 181 -SHSPSSCKPCDPCSPGSCGSGGEGENCOKLRIICAOQCSHRGGRSPSDCHQCAAG 239
QY 244 CTGPKHSDCLACILHFNHSGICELHCPALVTYNTDFEESPNFEGRYTGACVTCAPRY 303
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 240 CTGPRSDCLVQKQODEATCTCDTQCPPLMLYPTTYQMDVNEBGRKYSFGATCVKCKPRY 299
QY 304 LSTDVSGCTLVCPRLHNOEYTAEDGTQRCCKSCPKCAR 340
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 300 VVTDHSGCVRACGPDYEV-EDDGIRKCKKCDGPKR 335

RESULT 7
ERRB4_RAT STANDARD: PRT; 1308 AA.
AC 062956; 0922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbb-4 precursor (EC 2.7.1.112).
GN ERBB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnet P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
RN [4]
RP FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
RN 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
RN NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
RN NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

```



CC	tyrosine phosphatase.
CC	-I- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC	RECEPTORS (POTENTIAL).
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-I- ALTERNATIVE PRODUCTS: 2 isoforms; JM-A (SHOWN HERE) AND JM-B;
CC	ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC	FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC	BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC	NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC	CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC	-I- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC	KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC	PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC	LUNG, SALIVARY GLAND, AND PANCREAS.
CC	-I- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC	RESIDUES
CC	-I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sdb.ch/announc/">http://www.isb-sdb.ch/announc/</a>
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).
CC	-----
DR	EMBL; L07868; AAB59446.1; -
DR	HSSP; P11362; 1FGK.
DR	Genew; HGNC:3432; ERBB4.
DR	MIM: 600543; -
DR	InterPro; IPR000494; EGRF_L_domain.
DR	InterPro; IPR000719; Euk_kinase.
DR	InterPro; IPR002174; Furin-like.
DR	InterPro; IPR001245; Tyr_kinase.
DR	InterPro; IPR004019; YLP_motif.
DR	Pfam; PF00069; kinase; 1.
DR	Pfam; PF00757; Furin-like; 1.
DR	Pfam; PF01030; Recep_L_domain; 2.
DR	Pfam; PF02757; YLP; 2.
DR	ProDom; PD000001; Euk_kinase; 1.
DR	SMART; SM00261; FU; 4.
DR	SMART; SM00219; Tyrc; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW	transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW	transubstrate; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW	Alternative splicing.
KW	SIGNAL
FT	1 25 POTENTIAL.
FT	CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT	DOMAIN 26 651. EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 652 675 POTENTIAL.
FT	DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 186 334 CYS-RICH.
FT	DOMAIN 496 633 CYS-RICH.
FT	DOMAIN 718 985 PROTEIN KINASE.
FT	NP_BIND 724 732 ATP (BY SIMILARITY).
FT	BINDING 751 751 ATP (BY SIMILARITY).
FT	ACT_SITE 843 843 BY SIMILARITY.
FT	DIISULFID 189 197 BY SIMILARITY.
FT	DIISULFID 193 205 BY SIMILARITY.
FT	DIISULFID 213 221 BY SIMILARITY.
FT	DIISULFID 217 229 BY SIMILARITY.
FT	DIISULFID 230 238 BY SIMILARITY.
FT	DIISULFID 234 246 BY SIMILARITY.
FT	DIISULFID 249 258 BY SIMILARITY.
FT	DIISULFID 262 289 BY SIMILARITY.
FT	DIISULFID 293 304 BY SIMILARITY.
FT	DIISULFID 308 323 BY SIMILARITY.
FT	DIISULFID 326 330 BY SIMILARITY.
FT	DIISULFID 503 512 BY SIMILARITY.
FT	DIISULFID 507 520 BY SIMILARITY.
FT	DIISULFID 523 532 BY SIMILARITY.

Query Match	Best Local Similarity	Matches 157	Conservative 45	Mismatches 124	Indels 22	Gaps
FT DISULFID	536	552	BY SIMILARITY.			
FT DISULFID	555	569	BY SIMILARITY.			
FT DISULFID	559	577	BY SIMILARITY.			
FT DISULFID	580	589	BY SIMILARITY.			
FT DISULFID	593	614	BY SIMILARITY.			
FT DISULFID	617	625	BY SIMILARITY.			
FT DISULFID	621	633	BY SIMILARITY.			
FT MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).			
FT MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).			
FT MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).			
FT MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).			
FT CARBOHYD	138	138	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	174	174	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	181	181	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	253	253	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	358	358	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	410	410	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	473	473	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	495	495	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	548	548	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	576	576	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT CARBOHYD	620	620	N-LINKED (GLICNAC. . .) (POTENTIAL).			
FT VAMPPLIC	626	648	NGPTSHDCITYPTHTSHLPQHA -> IGSSTIEDICGLMD (IN ISOFORM JM-B)			
SO SEQUENCE	1308 AA	146807 MW	5E4AE80985D8761 CRC64			
Query Match	33.9%	Score 775	DB 1	Length 1308		
Best Local Similarity	45.1%	Pred. No. 2.3e-52				
Matches 157	Conservative 45	Mismatches 124	Indels 22	Gaps		
OY 9	WGLLALLPFGA----	STQVCTGTDMDKRLRASPETHLMDMLRLHYQSGOVYQGNIELTY	64			
DB 8	WVWVSLVLAAGVAPDSDSGVCGATENKISSLSDEQQYRALRKYYENCEVWGNLEITS		67			
OY 65	LPTNALSFLDIDIQVQGVLLAHNOVROVPLRLRYVGTQLEFEDNYVALAVLINDGDPIN	124				
DB 68	IEHNNDLSLRREVLTGVLVALNOFRLPLENRLIINGTLIEDRYALALFLNTRKGG	127				
OY 125	NTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPDLQCYQDTILMKDFPHKNNQALTLI	184				
DB 128	NF-----GLOELGKLNLEIILINGGVYVQNKFLCYADPIIHODIVRNWPMSNLTV	178				
OY 165	DINRSRACHPCSPMKSGSCWGESSEDDCSLRTVCAAGC-ARCKGRLPTDCHECCAG	243				
DB 179	STNGSSGGCGRCKSCTG-RCWGPTENHCOTLRTVCAEDCGDRCTGPRVSDCCHRCAG	237				
OY 244	CTGPKHSDCLALAHNHSIGICELHCPALTYNTDFEPMNPBEGRYTFGASCVTACPYNY	303				
DB 228	CSGPRGDTOCFAFMFNDSDGACVTGCPQRFVYVPTTFQLEHNNAKTYLCAFCVKKKPRHNF	297				
OY 304	ISTDVGSCITVCPRLHNOEYTAEDGTQRCCKSPCAR-----GHSIL	346				
DB 298	V-VDSSSCVACRPPSSKMEY-EDNGIKMKPCPDICPKACDGIGTGSLM	343				
RESULT 9						
ERR3_HUMAN						
ID ERR3_HUMAN	STANDARD:	PRT:	1342 AA			
AC P21860:						
DT 01-MAY-1991	(Rel. 18, Created)					
DT 01-MAY-1991	(Rel. 18, Last sequence update)					
DT 15-JUN-2002	(Rel. 41, Last annotation update)					
DE	Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)					
DE	(c-erbB3) (tyrosine kinase-type cell surface receptor HER3).					
GN	ERBB3 OR HER3.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=90083234; PubMed=2687875;					
RA	Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;					

RT "Isolation and characterization of ERBB3, a third member of the  
 RT ERBB/epidermal growth factor receptor family: evidence for  
 RT overexpression in a subset of human mammary tumors."  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE:90311312; PubMed:2164210;  
 RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,  
 RA Todaro G.J., Shoyab M.;  
 RT "Molecular cloning and expression of an additional epidermal growth  
 RT factor receptor-related gene."  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A. (SHORT FORM).  
 RC Tissue-Placenta;  
 RX MEDLINE:93282822; PubMed:7685162;  
 RA Katoh M., Yazaki Y., Sugimura T., Terada M.;  
 RT "c-erbB gene encodes secreted as well as transmembrane receptor  
 RT tyrosine kinase."  
 CC Biochem. Biophys. Res. Commun. 192:1189-1197(1993).  
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTAK.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND  
 CC SECRETED (SHORT FORM).  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A  
 CC SHORT SECRETED FORM. EXIST DUE TO ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.  
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 CC AND PROMOTES ITS ASSOCIATION WITH THE p65 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).  
 CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M29366; AAA35790.1; -  
 DR EMBL; M34309; AAA35979.1; -  
 DR EMBL; S61953; AAB26935.1; -  
 DR PIR; A36223; A36223.  
 DR HSSP; P1362; 1FGK.  
 DR GeneW; HGNC:3431; ERBB3.  
 DR MIM; 190151; -  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase.1.  
 DR Pfam; PF00757; Furin-like.1.  
 DR Pfam; PF01030; Recep\_L\_domain.2.  
 DR ProDom; PD000001; Euk\_pkinase.1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SMO0219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; FALSE\_NEG.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 DR Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Alternative splicing.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 1342 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.  
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 644 664 POTENTIAL.

FT	DOMAIN	665	1342	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	709	966	PROTEIN KINASE.
FT	NP_BIND	715	723	ATP (BY SIMILARITY).
FT	BINDING	742	742	ATP (BY SIMILARITY).
FT	ACT_SITE	834	834	BY SIMILARITY.
FT	DISULFID	186	194	BY SIMILARITY.
FT	DISULFID	190	202	BY SIMILARITY.
FT	DISULFID	210	218	BY SIMILARITY.
FT	DISULFID	214	226	BY SIMILARITY.
FT	DISULFID	227	235	BY SIMILARITY.
FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	552	565	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	VARSPLIC	141	183	ELISGVYKDKKCHMDTIDMDRIDVDRDAEIVKDNGR SC -> GQFPVPSGLTFPOGQDQVLLDDDDRLRLLSASSK VPTVLAAY (IN SHORT ISOFORM). MISSING (IN SHORT ISOFORM). E -> G (IN REF. 2). E -> G (IN REF. 2).
FT	VARSPLIC	184	1342	
FT	CONFLICT	560	560	
FT	CONFLICT	1064	1064	
SO	SEQUENCE	1342 AA;	148097 MW;	7201E766CA374BD CRC64;

Query Match 32.2%; Score 735.5; DB 1; Length 1342;  
 Best Local Similarity 44.0%; Pred. No. 2; 7e-49;  
 Matches 150; Conservative 49; Mismatches 121; Indels 21; Gaps 9;

QY	10	GILLALLPPGAA--STQVCTGTDMKRLRLPASPEHLDMRLHYGCGOVQGNLETLVLTPT	67
DB	11	GLLFSLARSEVGNQAVCPGTLNLGSLVYGDABENQYQTLIKLYERCEVYVMGLLEIVLIGH	70
QY	68	NASLSFLDIOEVGYVLIANQVROVPLQRLRIYRGTLQFEEDNYALAVLDNGDPLNNTT	127
DB	71	NADLSFLQWIREVGYVLIANNEFSTLPLNLRVVRGQVYDGKRAIFVM-----LVNVT	125
QY	128	PYTGASPGGLRQLRLSLTEILKGVLIQNRQOLCYQDTILMKDIFKKNQALFLIDTN	187
DB	126	---NSSHALRQLRLTQLTEILSGGVYIEKNDKLTCHMDTIDRDIVDRD---AAIVYKD	178
QY	188	RSRACHPSCPMCKSGSRCESSDOSTLRIVACAGG--ARCKGRLPTDCHQCAAGGCG	246
DB	179	NRSCPPCHVEYK--RCMGPSGEDQTLTKTICAPQCNCHCGPMPNOCCHDECAAGGCGS	237
QY	247	PKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNDEGRYTGASCAYTACPNYLTST	306
DB	238	PODTDFACRHFNDGACVPCRPQPLVYNNKLFQLEPMPHFYQYGVGVASCAPHNFV-V	296
QY	307	DVSGCTVCPHLNQVETAEADGQREKSKPCAR---GTHS	344
DB	297	DOTSCVRACPPDKMEVD--KNGIKMCEPGGGLCPKACEGTGS	336



```

RESULT 11
XMWK_XIPMA STANDARD: PRT: 1167 AA.
AC P13388:
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMWK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Witbrodt J., Adam D., Maltsechek B., Mauelel W., Raulf F.,
RA Telling A., Robertson S.M., Scharl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
RL inducing Ty locus in Xiphophorus.";
RL Nature 341:415-421(1989).
RN [2]
RP REVISION TO 515.
RA Scharl M.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16891; CAA34770.2; -.
CC PIR: S06142; S06142.
CC DR HSSP: P11362; 1FCG.
CC DR InterPro: IPR000494; EGFR_L.domain.
CC DR InterPro: IPR000719; Euk_Pkinase.
CC DR InterPro: IPR002174; Furin-like.
CC DR InterPro: IPR002290; Ser_thr_Pkinase.
CC DR InterPro: IPR001245; Tyr_Pkinase.
CC DR Pfam: PF00069; Pkinase; 1.
CC DR Pfam: PF00757; Furin-like; 1.
CC DR Pfam: PF01030; Recep_L.domain; 2.
CC DR PRINTS: PR00109; TYRKINASE.
CC DR ProDom: PD000001; Euk_Pkinase; 1.
CC DR SMART: SM00261; FU; 5.
CC DR SMART: SM00220; S_TKc; 1.
CC DR SMART: SM00219; TyKc; 1.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
CC
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
CC
CC SIGNAL 1 25
CC CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
CC KINASE.
CC DOMAIN 26 642 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 643 665 POTENTIAL.
CC DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 710 977 PROTEIN KINASE.
CC NP_BIND 716 724 ATP (BY SIMILARITY).
CC BINDING 743 743 ATP (BY SIMILARITY).
CC ACT_SITE 835 835 BY SIMILARITY.

```

```

FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 228 BY SIMILARITY.
FT DISULFID 224 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 311 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 333 337 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 524 533 BY SIMILARITY.
FT DISULFID 537 553 BY SIMILARITY.
FT DISULFID 536 569 BY SIMILARITY.
FT DISULFID 560 577 BY SIMILARITY.
FT DISULFID 593 615 BY SIMILARITY.
FT DISULFID 618 626 BY SIMILARITY.
FT DISULFID 622 634 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DCID55A CRC64;

```

```

Query Match 29.8%; Score 681.5; DB 1; Length 1167;
Best Local Similarity 42.4%; Pred. No. 3; Se=45;
Matches 145; Conservative 46; Mismatches 136; Indels 15; Gaps 8;

```

```

QY 4 AALCRGLALLLPAAST---QVCTGDMKRLPASPEFLDMLRHLYGCGVYVGN 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8 AALLQ--LLLVLSISCCSTDPKRCQGSNMTM---LDNHYLKKMKKYSQCNVLE 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 60 LEITYPTNNSLSEFLDIDIEVOGYVLIHNOYRQVPLRLIYRGTLFEDNYALAVLN 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 63 LEITYQENODLSFLQSIQIEVGGYVLIANNEVSTIPLVNLRLIRGONLYEGNFTLLVMSN 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 120 GDPNLTFTVYTGASPGGLAELOLRSTELLKGVLLQRRPOLCYOTFIIMKDFHKNOL 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 YOK-NPSSP--DVYQYGLQLOLSNLTETLLSGVKVSHNPPLCNVETIMWMDIVDTSNP 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 180 ALTLIDITNSRACHPCSPMKCKSGWSESDQSLTFRYCAGGC-ARCKGPLPTCCHE 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 180 TMLLIHAFERQCKDKDHGCVNGSCNAPRGHCQKTKILCAEQCKRRRGKRPIDCNE 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 239 QCAAGCTGPKHSDCLALFHNSGICELHCPALVYNTDTFESMPNDEGRYTFGASCVA 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 240 HCAGCGTGRATDCIACRDFNDDGTCKDPPKIDIVSHQVNDPNIKYFGACVKE 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 299 CPYNYISTDVGSLVCPRLHNEVYTAEDGTQRCCKSKCAR 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 300 CPSNVYVTE-GACVRSASGMEVD-ENGRSKRCPCDGVCPK 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 12
EGFR_DROME STANDARD: PRT: 1426 AA.
AC P04412; 061601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

```

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RX MEDLINE=94350209; PubMed=8070664;  
 RA Clifford R., Schupbach T.;  
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals  
 RT subdomains of the receptor protein.";  
 RL Genetics 137:531-550(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Clifford R., Schupbach T.;  
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RX MEDLINE=85124611; PubMed=2982499;  
 RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;  
 RT "The Drosophila EGF receptor gene homolog: conservation of both  
 RT hormone binding and kinase domains.";  
 RL Cell 40:599-607(1985).  
 RN [4]  
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
 RC STRAIN=Oregon-R; TISSUE=Embryo;  
 RX MEDLINE=87002474; PubMed=3093080;  
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;  
 RT "Alternative 5' exons and tissue-specific expression of the  
 RT Drosophila EGF receptor homolog transcripts.";  
 RL Cell 46:1091-1101(1986).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=99102120; PubMed=9882502;  
 RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;  
 RT "Several levels of EGF receptor signaling during photoreceptor  
 RT specification in wild-type, Ellipse, and null mutant Drosophila.";  
 RL Dev. Biol. 205:139-144(1999).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champé M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Abrell J.F., Adbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bessie P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
 RA Borovda D., Botchan M.A., Boulter J., Brokstein P., Brotler P.,  
 RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,  
 RA Foster C.A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwu C.,  
 RA Jaitani B.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Modarres C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relaut K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong E.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [7]  
 RP SEQUENCE OF 959-1078 FROM N.A.  
 RC STRAIN=Daekwanryeong;  
 RX MEDLINE=85137938; PubMed=2983232;  
 RA Wadsworth S.C., Vincent W.S. III, Blodeau-Wentworth D.;  
 RT "A Drosophila genomic sequence with homology to human epidermal  
 RT growth factor receptor.";  
 RL Nature 314:178-180(1985).  
 RN [8]  
 RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=92038942; PubMed=1936959;  
 RA Raz E., Schejter E.D., Shilo B.-Z.;  
 RT "Interallelic complementation among DER/Elb alleles: implications for  
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";  
 RL Genetics 129:191-201(1991).  
 RN [9]  
 RP REVIEW.  
 RX MEDLINE=97248481; PubMed=9094709;  
 RA Perrimon N., Perkins L.A.;  
 RT "There must be 50 ways to rule the signal: the case of the Drosophila  
 RT EGF receptor.";  
 RL Cell 89:13-16(1997).  
 RN [10]  
 RP FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,  
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-  
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.  
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES. AND FOR THE  
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE  
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOTEROSA  
 CC AND VENTRAL NEUROECODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
 CC CUTICLE.  
 CC -i- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein  
 CC tyrosine phosphate.  
 CC -i- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
 CC PROTEIN.  
 CC -i- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND  
 CC TYPE III: ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -i- TISSUE SPECIFICITY: UNBOUTHOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,  
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANTENNA OF  
 CC TESTIS AND OVARY. AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH  
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
 CC AND THORACIC AND ABDOMINAL GANGLIA.  
 CC -i- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -i- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF052754; AAC08536.1; -  
 DR EMBL: AF052753; AAC08536.1; JOINED.  
 DR EMBL: AF052754; AAC08535.1; -  
 DR EMBL: AF052752; AAC08535.1; JOINED.  
 DR EMBL: K03054; AAA51462.1; -  
 DR EMBL: K03417; AAA51460.1; -  
 DR EMBL: K03416; AAA50965.1; -  
 DR EMBL: K03418; AAA51461.1; -  
 DR EMBL: AF109077; AAD26134.1; -



```

DR EMBL: AF109078; AAD26132.1; -.
DR EMBL: AF109082; AAD26132.1; JOINED.
DR EMBL: AF109078; AAD26133.1; -.
DR EMBL: AF109084; AAD26133.1; JOINED.
DR EMBL: AF109079; AAD26130.1; -.
DR EMBL: AF109081; AAD26130.1; JOINED.
DR EMBL: AF109079; AAD26131.1; -.
DR EMBL: AF109083; AAD26131.1; JOINED.
DR EMBL: AF109080; AAD26131.1; -.
DR EMBL: AF003454; AAF46732.1; -.
DR EMBL: X02293; CAA26157.1; -.
DR EMBL: X78920; CAA55523.1; -.
DR EMBL: X78918; CAA55521.1; -.
DR EMBL: X78919; CAA55522.1; -.
DR PIR: A00640; GOFE.
DR HSSP: P11362; 1FGK.
DR Flybase: FBgn0003731; Egfr.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; kinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00261; FU; 7.
DR SMART: SM00219; TyrcK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane: Glycoprotein; Receptor; Phosphorylation; Transferase;
KM Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW developmental protein.
FT SIGNAL 1 30
FT CHAIN 31 1426
FT DOMAIN 31 868
FT TRANSMEM 869 889
FT DOMAIN 890 1426
FT DOMAIN 938 1198
FT NP_BIND 944 952
FT BINDING 971 971
FT ACT_SITE 1063 1063
FT MOD_RES 902 902
  POTENTIAL.
  EPIDERMAL GROWTH FACTOR RECEPTOR.
  EXTRACELLULAR (POTENTIAL).
  POTENTIAL.
  CYTOPLASMIC (POTENTIAL).
  PROTEIN KINASE.
  ATP (BY SIMILARITY).
  ATP (BY SIMILARITY).
  BY SIMILARITY.
  PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
Query Match 25.1%; Score 574.5; DB 1; Length 1426;
Best Local Similarity 36.3%; Pred. No. 8.8e-37;
Matches 118; Conservative 46; Mismatches 130; Indels 31; Gaps 7;

```

```

RESULT 13
LET23_CAEL
ID LET23_CAEL STANDARD; PRT; 1323 AA.
AC P24348;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Let-23 receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN LET-23 OR KIN-7 OR ZK1067.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91080919; PubMed=1979659;
RA Aroian R.V., Koga M., Mendel J.E., Ohshima Y., Sternberg P.W.;
RT "The let-23 gene necessary for Caenorhabditis elegans vulval
  induction encodes a tyrosine kinase of the EGF receptor subfamily.";
RL Nature 348:693-699(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2; PubMed=8604137;
RA Sakai T., Koga M., Ohshima Y.;
RT "Genomic structure and 5' regulatory regions of the let-23 gene in
  the nematode C. elegans.";
RL J. Mol. Biol. 256:548-555(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Thomas K.;
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RL [4]
RP MUTANTS.
RA MEDLINE=94147981; PubMed=8313880;
RA Aroian R.V., Les G.M., Sternberg P.W.;
RT "Mutations in the Caenorhabditis elegans let-23 EGFR-like gene define
  elements important for cell-type specificity and function.";
RL EMBO J. 13:360-366(1994).
RN [5]
RP SUBCELLULAR LOCATION.
RA STRAIN=Bristol N2;
RA MEDLINE=99287744; PubMed=10359617;
RA Whitfield C.W., Benard C., Barnes T., Hekimi S., Kim S.K.;
RT "Basolateral localization of the Caenorhabditis elegans epidermal
  growth factor receptor in epithelial cells by the PDZ protein
  LIN-10.";
RL Mol. Biol. Cell 10:2087-2100(1999).
CC -!- FUNCTION: Tyrosine kinase receptor required for the induction of
  vulval differentiation. Possible receptor for the inductive signal
  required for vulval development. Activated by lin-3 and acts by
  way of let-60 RAS. The lin-3/let-23 pair is a simplified version
  of the mammalian neuroligin-ERBB network.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein
  tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Basolateral and
  apical membrane of cell junctions in epithelial vulval precursor
  cells.
CC -!- TISSUE SPECIFICITY: Vulval precursor cells.
CC -!- DEVELOPMENTAL STAGE: Expressed during L2 and L3 larval stages.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC EMBL: X57767; CAA40919.1; ALT_SEQ.

```



```
DR HSSP: P06213; 11RK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003962; FNIII_repeat.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002011; RTK_kinaseII.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00041; fn3; 3.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recept_L-domain; 2.
DR PRINTS: PR00014; FNTPEI11.
DR PRINTS: PR00109; TYRKINASE.
DR PRODOM: PD000001; Euk_pkinase; 1.
DR SMART: SM00060; FN3; 3.
DR SMART: SM00261; FU; 1.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KM Transferrase: Tyrosine-protein kinase: Receptor; Transmembrane;
KM Glycoprotein: ATP-binding; Phosphorylation; Signal.
FT SIGNAL 1 29
FT CHAIN 30 716
FT PROPEP 717 720
FT CHAIN 721 1363
FT DOMAIN 721 928
FT TRANSMEM 929 949
FT DOMAIN 950 1363
FT DOMAIN 994 1283
FT NP_BIND 1000 1008
FT BINDING 1028 1028
FT ACT_SITE 1148 1148
FT MOD_RES 1174 1174
FT CARBOHYD 51 51
FT CARBOHYD 97 97
FT CARBOHYD 137 137
FT CARBOHYD 278 278
FT CARBOHYD 483 483
FT CARBOHYD 599 599
FT CARBOHYD 617 617
FT CARBOHYD 665 665
FT CARBOHYD 666 666
FT CARBOHYD 711 711
FT CARBOHYD 732 732
FT CARBOHYD 736 736
FT CARBOHYD 743 743
FT CARBOHYD 816 816
FT CARBOHYD 885 885
FT CARBOHYD 898 898
SQ SEQUENCE 1363 AA; 154104 MW; 238120B4EAB1ED65 CCR64;

Query Match 15.0%; Score 342.5; DB 1; Length 1363;
Best Local Similarity 28.3%; Pred. No. 8e-19;
Matches 106; Conservative 40; Mismatches 124; Indels 105; Gaps 20;
```

```
DB 157 IDMSFIAESGYSNN-----FIVDNREEECVAFCCGRCRIKHPLVLDICMAE--EHCQKVC 210
OY 217 RTVAGGCARCKRGLPDDCCHECAACCTGPKHSDCIACLHFNHSGICELHCPALVYNT 276
DB 211 PESLGNCR---DGISGCCHEHCIGCGDPTERDVCACRYFVNHSGCLLCQPDYQYK 266
OY 277 D---TFESMPNPEGRY--TFGASCVCAPYNYSTVGS---CTLYCPLHNGEVTAEAG 327
DB 267 DRCITFEECPCNTNSWVKLHHRKCIPECFSGY--TTDINNRLCT----- 310
OY 328 TQCE-KCSRPCARG 341
DB 311 --EEGCGCKRCKG 323

RESULT 15
INSR_DROME
ID INSR_DROME STANDARD; PRT; 2146 AA.
AC P09208; Q24089; Q24023;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Insulin-like receptor precursor (EC 2.7.1.112).
GN INR OR INR-A OR DIR-A.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95354555; PubMed=7628438;
RA Fernandez R., Tabarini D., Azpiroz N., Frasch M., Schlessinger J.;
RT "The Drosophila insulin receptor homolog: a gene essential for
RT embryonic development encodes two receptor isoforms with different
RT signaling potential."
RL J. Biol. Chem. 270:4236-4243(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95181404; PubMed=7876183;
RA Ruan Y., Chen C., Cao Y., Garofalo R.S.;
RT "The Drosophila insulin receptor contains a novel carboxyl-terminal
RT extension likely to play an important role in signal transduction."
RL J. Biol. Chem. 270:4236-4243(1995).
RN [3]
RP SEQUENCE OF 652-1749 FROM N.A.
RX STRAIN=Drosophila; TISSUE=Embryo;
RC MEDLINE=87100165; PubMed=3099787;
RA Nishida Y., Hata M., Nishizuka Y., Rutter W.J., Ebina Y.;
RT "Cloning of a Drosophila cDNA encoding a polypeptide similar to the
RT human insulin receptor precursor."
RL Biochem. Biophys. Res. Commun. 141:474-481(1986).
RN [4]
RP SEQUENCE OF 1297-1595 FROM N.A.
RX MEDLINE=86259667; PubMed=3014506;
RA Petruzzelli L., Herrera R., Arenas-Garcia R., Fernandez R.,
RA Birnbaum M.J., Rosen O.M.;
RT "Isolation of a Drosophila genomic sequence homologous to the kinase
RT domain of the human insulin receptor and detection of the
RT phosphorylated Drosophila receptor with an anti-peptide antibody."
RL Proc. Natl. Acad. Sci. U.S.A. 83:4710-4714(1986).
CC -1- FUNCTION: THIS RECEPTOR PROBABLY BINDS AN INSULIN RELATED PROTEIN
CC AND HAS A TYROSINE-PROTEIN KINASE ACTIVITY (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY DISULFIDE
CC BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF THE LIGAND-
CC BINDING DOMAIN, WHILE THE BETA CHAINS CARRY THE KINASE DOMAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC RECEPTOR SUBFAMILY.
```

[illegible][illegible]

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:31:26 ; Search time 46.2751 Seconds  
(without alignments)  
1865.663 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287  
Sequence: 1 METALCRWGLLALLPPGA.....YGRGPPDAHVAVLNLSRYEG 419

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_21: \*  
1: sp.\_archaea: \*  
2: sp.\_bacteria: \*  
3: sp.\_fungi: \*  
4: sp.\_human: \*  
5: sp.\_invertebrate: \*  
6: sp.\_mammal: \*  
7: sp.\_mhc: \*  
8: sp.\_organelle: \*  
9: sp.\_phage: \*  
10: sp.\_plant: \*  
11: sp.\_rodent: \*  
12: sp.\_virus: \*  
13: sp.\_vertebrate: \*  
14: sp.\_unclassified: \*  
15: sp.\_virus: \*  
16: sp.\_bacteria: \*  
17: sp.\_archae: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2275	99.5	419	4	Q9UK79
2	1716.5	75.1	1259	6	O18735
3	794	34.7	527	13	Q90836
4	789	34.5	643	11	Q9ERV6
5	789	34.5	655	11	Q9WVF5
6	789	34.5	1210	11	Q9EP98
7	781.5	34.2	1209	11	Q9QX70
8	773	33.8	478	11	Q9SE50
9	734	32.1	331	4	Q9BUD7
10	723	31.6	149	6	Q9BG66
11	693	30.3	1165	13	Q9YH40
12	662.5	29.0	1328	13	P79754
13	647	28.3	599	13	Q9PSH2
14	571.5	25.0	1433	5	Q9B1H9
15	463.5	20.3	1137	13	Q9W6F6
16	404.5	17.7	150	6	Q9BG64

17	402.5	17.6	1368	5	Q23821	Q23821 caenorhabd
18	395	17.3	151	6	Q9BG65	Q9bg65 oryctolagus
19	363.5	15.9	366	5	Q26569	Q26569 schistosoma
20	363.5	15.9	1717	5	Q26566	Q26566 schistosoma
21	353.5	15.5	1193	5	Q9Y1X8	Q9y1x8 ephydalia f
22	331	14.5	334	5	Q26567	Q26567 schistosoma
23	331	14.5	342	5	Q26568	Q26568 schistosoma
24	300.5	13.1	1418	13	Q8UW83	Q8uw83 paralichthy
25	300	13.1	1472	5	Q9U5A8	Q9u5a8 bombyx mori
26	298.5	13.1	1671	5	Q9N1V5	Q9n1v5 biophthalmi
27	298	13.0	1358	13	Q73798	Q73798 xenopus lae
28	298	13.0	1418	13	Q93457	Q93457 scophthalmu
29	294	12.9	2144	5	Q9VD94	Q9vd94 drosophila
30	284	12.4	1362	13	Q9PVZ4	Q9pvz4 xenopus lae
31	280	12.2	1749	5	Q8T0W6	Q8t0w6 echinococcu
32	278	12.2	1412	13	Q8UW84	Q8uw84 paralichthy
33	275.5	12.0	1369	13	Q8UW86	Q8uw86 paralichthy
34	273.5	12.0	89	11	Q88459	Q88459 mus musculu
35	272	11.9	469	11	Q63721	Q63721 rattus norv
36	270	11.8	410	11	Q63720	Q63720 rattus norv
37	256.5	11.2	1368	13	Q8UW85	Q8uw85 paralichthy
38	252.5	11.0	1371	11	Q9QVW4	Q9qv4 rattus sp.
39	232	10.1	1245	13	Q9YGH8	Q9ygh8 scophthalmu
40	230	10.1	946	5	Q9YJ04	Q9yj04 drosophila
41	210	9.2	868	5	Q9VEE2	Q9vee2 drosophila
42	203	8.9	1843	5	Q968Y9	Q968y9 caenorhabd
43	203	8.9	1846	5	Q16131	Q16131 caenorhabd
44	191	8.4	82	6	Q9NOK4	Q9nok4 sus scrofa
45	189	8.3	131	5	Q9BH16	Q9bh16 anopheles g

## ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	419 AA.
Q9UK79			
AC Q9UK79:			
DT 01-MAY-2000 (TREMBLrel. 13, Created)			
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE Herstatin.			
GN HER-2.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=99415951; PubMed=10485918;			
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;			
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted			
RT autolnhibitor.";			
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).			
RN [2]			
RP SEQUENCE FROM N.A.			
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;			
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.			
DR EMBL; AF177761; AAD56009.2; -			
DR InterPro: IPR000494; EGFR_L-domain.			
DR InterPro: IPR002174; Furin-like.			
DR Pfam: PF00757; Furin-like: 1.			
DR Pfam: PF01030; Recep_L-domain: 1.			
DR SMART; SM00261; FU; 1.			
SQ SEQUENCE 419 AA: 45472 MW; FECLBE347E2D030C CRC64;			
Query Match	99.5%;	Score 2275;	DB 4; Length 419;
Best Local Similarity	99.5%;	Pred. No. 3e-200;	
Matches 417;	Conservative 1;	Mismatches 1;	Indels 0; Gaps 0;
QY 1 METALCRWGLLALLPPGAATQVCTGTDMLRLPASPETHLMDLRHLVYGGQVVGML 60			
DB 1 METALCRWGLLALLPPGAATQVCTGTDMLRLPASPETHLMDLRHLVYGGQVVGML 60			

```
QY 61 ELTYLPTNASTSLFLODIOEVGYVLIHAHQVROVPLQRLRIVRGTQLEFEDNVALAVLNDG 120
    |||||
DB 61 ELTYLPTNASTSLFLODIOEVGYVLIHAHQVROVPLQRLRIVRGTQLEFEDNVALAVLNDG 120
QY 121 DPLNNTTPVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
    |||||
DB 121 DPLNNTTPVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGSRCKWGESSEDCSLTRTYVAGGCARCKGLPTDCCHEOC 240
    |||||
DB 181 LTLIDNRSRACHPCSPMCKSGSRCKWGESSEDCSLTRTYVAGGCARCKGLPTDCCHEOC 240
QY 241 AACCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRYTFGASCTYACP 300
    |||||
DB 241 AACCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRYTFGASCTYACP 300
QY 301 YNLTSDVSGSCTLVCPLNQEVTAEDGTORCEKSKPCARGTSLPRPAVVPPLRMOP 360
    |||||
DB 301 YNLTSDVSGSCTLVCPLNQEVTAEDGTORCEKSKPCARGTSLPRPAVVPPLRMOP 360
QY 361 GPAHPVLSFLRPSMDLVSAFYSLPLAPLSTVPIPSVYGRGPDPAHYAVNLSRREG 419
    |||||
DB 361 GPAHPVLSFLRPSMDLVSAFYSLPLAPLSTVPIPSVYGRGPDPAHYAVNLSRREG 419
```

## RESULT 2

```
018735 PRELIMINARY: PRT: 1259 AA.
AC 018735;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erbb-2
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RT "CDNA cloning of erbb-2 from canine mammary gland."
RT Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB008451; BAA23127.1;
DR HSSP: P11362; 1FGK.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PR00757; Furin-like.1.
DR Pfam: PR00069; pkinase.1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FY; 3.
DR SMART: SM00219; TyrKc; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ARP; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364DA9C4ACD6 CRC64;
```

Query Match 75.1%; Score 1716.5; DB 6; Length 1259;  
Best Local Similarity 82.2%; Pred. No. 2e-148;  
Matches 324; Conservative 14; Mismatches 49; Indels 7; Gaps 2;

```
QY 1 MELAALCRNGLLALLPPGASLTQVCTGTDMLRLPASPTHTDMLRHLYOGCGVYQGNL 60
    |||||
DB 1 MELAALCRNGLLALLPPGASLTQVCTGTDMLRLPASPTHTDMLRHLYOGCGVYQGNL 60
```

```
QY 61 ELTYLPTNASTSLFLODIOEVGYVLIHAHQVROVPLQRLRIVRGTQLEFEDNVALAVLNDG 120
    |||||
DB 61 ELTYLPTNASTSLFLODIOEVGYVLIHAHQVROVPLQRLRIVRGTQLEFEDNVALAVLNDG 120
QY 121 DPLNNTTPVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
    |||||
DB 121 DPLNNTTPVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGSRCKWGESSEDCSLTRTYVAGGCARCKGLPTDCCHEOC 240
    |||||
DB 181 LTLIDNRSRACHPCSPMCKSGSRCKWGESSEDCSLTRTYVAGGCARCKGLPTDCCHEOC 240
QY 241 AACCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRYTFGASCTYACP 300
    |||||
DB 241 AACCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRYTFGASCTYACP 300
QY 301 YNLTSDVSGSCTLVCPLNQEVTAEDGTORCEKSKPCAR-----GTHSLPRPAVVP 355
    |||||
DB 301 YNLTSDVSGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCGYGLMEHLREYRAVTSAN 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLPLAP 387
    |||||
DB 361 IQEFAAGCKKIFGSLAPLPESFDGDPASNTAPLP 394
```

## RESULT 3

```
090836 PRELIMINARY: PRT: 527 AA.
AC 090836;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE EGF/EGF-alpha receptor precursor.
GN C-ERAB.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE-92123214; PubMed-1732751;
RT Flickinger T.W., Mahle N.J., Kung H.-J.;
RT "An alternatively processed mRNA from the avian c-erbB gene encodes a
RT soluble, truncated form of the receptor that can block ligand-
RT dependent transformation."
RT Mol. Cell. Biol. 12:883-893(1992).
RL EMBL: M77637; AAA48759.1;
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PR00757; Furin-like.1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FY; 2.
DR SMART: SM00261; FY; 2.
FT SIGNAL 1
FT CHAIN 29 527
FT TAG 1 28
FT SIGNAL 29 527
SQ SEQUENCE 527 AA; 58353 MW; 7645644ABC095298 CRC64;
```

Query Match 34.7%; Score 794; DB 13; Length 527;  
Best Local Similarity 46.0%; Pred. No. 2.7e-64;  
Matches 155; Conservative 55; Mismatches 109; Indels 18; Gaps 7;

```
QY 11 LILALLPPGASLTQVCTGTDMLRLPASPTHTDMLRHLYOGCGVYQGNLETLVL 65
    |||||
DB 20 LILALLPPGASLTQVCTGTDMLRLPASPTHTDMLRHLYOGCGVYQGNLETLVL 65
QY 66 PTNASTSLFLODIOEVGYVLIHAHQVROVPLQRLRIVRGTQLEFEDNVALAVLNDG 125
    |||||
DB 80 EHRDLTFLKTIQEVAGVYLIHAHQVROVPLQRLRIVRGTQLEFEDNVALAVLNDG 138
QY 126 TTPVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 185
    |||||
DB 139 TQ-----GRELPMKRSLTEILNGVKISNPKLCNMDTVLMDIIDSRSK-PLTVLD 189
```

QY	186	-TNSRACHPSPKCSGRCGESEEDOSLTPVVCAGGA - RCKGPLPDDCHEOCAG	243
	190 <td>FASNUSSCPKRNPCETEDHCAGAEQMCQTILTKYCAQOCSGRCRGVVPDDCCNQCAG<td>249</td></td>	FASNUSSCPKRNPCETEDHCAGAEQMCQTILTKYCAQOCSGRCRGVVPDDCCNQCAG <td>249</td>	249
QY	244	CTGPKHSDCLACLFHNSHGICELHCAPALVTYNTDTEESMPNPEGRTYFGASCYACAPYNT	3033
Db	250	CTGGRSDCLACRFRFDDATCCKDTCPCPLVLYNPTTYQMDVNEPEKYSFGATVRECPHNY	309
QY	304	LSTDVSGCTLVCPRLHNOEYVTAEDTQCEKSCRKAC	340
Db	310	VVTHGSCVRSRCNTDYEV -EENGVRCKRCKDGLCSK	345
RESULT 4			
QSERV6	QSERV6	PRELIMINARY;	PRT; 643 AA.
AC	QSERV6;		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Epidermal growth factor receptor isoform 2.		
GN	EGFR.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STPAIN-C3H/101, 129/SVJ, AND 129/SVEVTPAC;		
RA	Relter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,		
RA	Shichair C.S., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,		
RA	Balsabrumblam S., Crossley T.O., Magnuson T.R., James C.D.,		
RA	Malhe N.J.,		
RT	"Comparative genomic sequence analysis and isolation of human and		
RT	mouse alternative Egr1 transcripts encoding truncated receptor		
RT	isoforms."		
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF275366; AAC28046.1; -		
DR	EMBL: AF275364; AAC28046.1; JOINED.		
DR	EMBL: AF275365; AAC28046.1; JOINED.		
DR	MGI: 95294; Egr1.		
DR	InterPro: IPR000345; GYTC_heme_bind.		
DR	InterPro: IPR000494; EGFR_L_domain.		
DR	InterPro: IPR002174; Furin-like.		
DR	Pfam: PF00757; Furin-like; 1.		
DR	Pfam: PF01030; Recep_L_domain; 2.		
DR	SMART: SM00261; FU; 4.		
DR	PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.		
KW	Receptor.		
SO	SEQUENCE		
	643 AA; 71476 MW; DEF22002C84911B1 CRC64;		
Query Match			
	34.5%; Score 789; DB 11; Length 643;		
	Best Local Similarity 46.3%; Pred. No. 9, 8e-64;		
	Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps		
QY	11	LLALLPPGAA--STQCTGDMKRLRAPERHLDMLRHLHLCQVQVOCVQVQVCELTETLYPT 68	
Db	14	LITRLACAGALEEKKVCQGSNRLTDLGTFEDHFLSLQMTYNNCEVVLGNLEITTYQRN 73	
QY	69	ASLSFLDIDDEVQGVLLAHNOYRQVPLQRLRIYRGTLQFEDYVALAVLDNGPLNTPP 128	
Db	74	YDLSEFLTIOEVAGVYLIALNTVERIPLENQIIRGNALYENTYVALAISN----- 124	
QY	129	VTGASPGGIRLQRLSLTELKAGVLLQRRPOLCYOTIIMKDI---FKKNQALATLI 184	
Db	125	-YGNRPGLRELPRNLQDEILIGAVRESNPPLICMDPTQMRDIQVONFVSNMSMDL--- 180	
QY	185	DTNSRACHPSPKCSGRCGESEEDOSLTPVVCAGGA - RCKGPLPDDCHEOCAG 243	
Db	181	-QSHRSPCPKRDPCSRGSCWGGEGEENQCKLTKICQAQCSHRCRGSPDCCNQCAG 239	
QY	244	CTGPKHSDCLACLFHNSHGICELHCAPALVTYNTDTEESMPNPEGRTYFGASCYACAPYNT 3033	

**RESULT 5**

ID	Q9WVF5	PRELIMINARY:	PRT;	655 AA.
AC	Q9WVF5;			
DT	01-NOV-1999 (TrEMBLrel. 12, Created)			
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)			
DI	01-JUN-2002 (TrEMBLrel. 21, last annotation update)			
DE	Epidermal growth factor receptor (epidermal growth factor receptor isoform 3).			
GN	Egfr.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRFAIN-C57BL/6J; TISSUE=LIVER;			
RA	Reller J.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Malhe N.J.;			
RT	"Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors."			
RL	Submitted (JAN-1999) to the EMBL/GenBank/DDB databases.			
RN	[2]			
RC	SEQUENCE FROM N.A.			
RC	STRFAIN-C3H/101, 129/SVJ, AND 129/SVEVJAC;			
RA	Reller J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearisall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Malhe N.J.;			
RT	"Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms."			
RL	Submitted (JUN-2000) to the EMBL/GenBank/DDB databases.			
RP	SEQUENCE FROM N.A.			
RC	STRFAIN-C57BL/6J; TISSUE=LIVER;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi T., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiysawa H., Kondo S., Yamataka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Glassi C., King B., Kochiya H., Knehl P., Lewis S., Matsuo Y., Nikido I., Pesole G., Quackenbush J., Schirral L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Bojelli D., Bojtunga N., Carninci P., de Bonaldo M.F., Brownstein W.J., Bolt C., Fletcher C., Fujita M., Gariboldi M., Gastunicich S., Hill D., Hofman M., Hume D.A., Kamuya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarella J., Mombeerts P., Nordore P., Ring B., Schoenbach C., Rodriguez I., Sakamoto N., Sasaki K., Sato K., Schenkeveld M., Seya T., Shibata Y., Storck K.-F., Suzuki H., Toyokura K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S., Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001).			
DR	EMBL; AF124513; AAD44149.1; -			
DR	EMBL; AF275366; AAG28047.1; -			
DR	EMBL; AF275364; AAG28047.1; JOINED.			
DR	EMBL; AF275365; AAG28047.1; JOINED.			
DR	EMBL; AK004944; BAB23688.1; -			
DR	EMBL; AK004883; BAB23641.1; -			
DR	EMBL; AK004911; BAB23662.1; -			





RA Pelech L.V., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,  
RA Earp H.S.;  
RT "A truncated, secreted form of the epidermal growth factor receptor is  
RT encoded by an alternatively spliced transcript in normal rat tissue";  
RL Mol. Cell. Biol. 10:2973-2982(1990).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-FISHER; TISSUE=LIVER;  
RC Pelech L.A.;  
RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.  
RP STRAIN-FISHER; TISSUE=LIVER;  
RA Guttridge K., Dawson T.L., Earp H.S.;  
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
RN EMBL: M31369; AAF14008.1. -  
DR HSP: p11362; 1PKK.  
DR InterPro: IPR000494; EGFR\_Ldomain.  
DR InterPro: IPR000719; Euk\_p\_kinase.  
DR InterPro: IPR001245; Furin-like.  
DR InterPro: IPR001245; Tyr\_kinase.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF00069; kinase; 1.  
DR Pfam: PF01030; Recep\_Ldomain; 2.  
DR PRINTS; PRO0109; TYRKINASE.  
DR ProDom: PD000001; Euk\_p\_kinase; 1.  
DR SMART; SM00261; FUK; 3.  
DR SMART; SM00219; TYRK; 1.  
DR PROSITE: PSS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PSS0011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PSS0109; PROTEIN\_KINASE\_TYR; 1.  
DR ATP-binding; Receptor; Transducer; Tyrosine-protein kinase.  
SQ SEQUENCE 1209 AA: 134891 MW: 96FEETFECCIB773 CRC64;

Query Match	34.2%	Score 781.5;	DB 11;	Length 1209;
Best Local Similarity	46.0%;	Pred. NO. 1.1e-62;		
Matches 157; Conservative	44;	Mismatches 117;	Indels 23;	Gaps 7;

[illegible]

RESULT	Q9E9E0	Q9E9E0
8	Q9E9E0	Q9E9E0
ID	Q9E9E0	PRELIMINARY; PRT; 478 AA.
AC	Q9E9E0;	
DT	01-MAR-2001 (Tremblrel. 16, Created)	
DT	01-OCT-2001 (Tremblrel. 18, Last sequence update)	
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)	
DE	Epidermal growth factor receptor related protein.	
GN	ERKP.	
OS	Rattus norvegicus (Rat).	

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=GASTRO-DUODENAL, MUCOSUS;  
RA Yu Y., Moshier J.A., Majumdar A.P.N.;  
RT "Cloning of a novel EGR-related peptide: A putative negative  
RT regulator of EGR."; [PMID:10560101](#)  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF187818; AAC17037.2; -  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR002174; Furin-like.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Receptor.  
KW SEQUENCE.  
SQ 478 AA; 5323 MW; CF873A8376C519P5 CRC64;

Query Match	33.8%;	Score 773;	DB 11;	Length 478;
Best Local Similarity	45.7%;	Pred. No. 2e-62;		
Matches 154;	Conservative 43;	Mismatches 118;	Indels 22;	Gaps 6

QY	11	LLALLLPPGAA--STQVCTGDMKLRLPAPSETHLIDLRLHLYGCGVQGNLETTLYPTN	68
Db	14	LLPFLAAGAAGLLEBKVKVCGQGSNNRLQLDGRFEHFLSLQGMNMCNCEVVLNLETTLYQNR	73
QY	69	ASLSFLDIDIGEVGYVLIANHNVQVFLQRLRYRGQLGFEDNLTALAVLNDGRLPNTTP	128
Db	74	YDLSLELTIEWAVGYFLIALNLTWVRIPSEDDQIIRGALYENYVLAALSEN-----	124
QY	129	VYSGAPGRLRELQRLSTLEILKGVLLFQRPNDQLCYDTIYMKDI---FKNNQIATLLI	184
Db	125	-YGINRGLRELRLPRNLQELLIGAVRSPNNPILCNMDITQIMRDLYGVNFVSNMSMDL---	180
QY	185	DTNSRACHPCSPMKCKSRCKGSESECCQSILTRVCAAGGA-RCKGRLPTDCCHEGCAAG	243
Db	181	-QSPSPSCPKCDPSCPGSCGSGCGGEECKQKLTIKICAQOQSHRCGRSPSDCCNCCAAAG	239
QY	244	CTGGRHSDCLACLTFNHSIGCIELCPALVYNTDTEFSMPNEGRGYTFGASCYACAPYNN	303
Db	240	CTGPGKSDCLVCGQKQFQEDATCKDTCPLMLYNPTTYGMNVNPEKYSFGATCVKNCPRNY	299
QY	304	LSTVSGSCTLVCPHLNDEVIAEDGTQCEKCSKPCAR	340
Db	300	VYTDHSGCVRACGPDYVEE-EEDGIRKCKCKDDGCRK	335

RESULT	9
09BUD7	
ID	09BUD7
AC	PRELIMINARY;
DT	01-JUN-2001 (TREMBLrel. 17, Created)
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Similar to v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=PLACENTA;
RA	Strausberg R.;
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL: BC002706; AAH02706.1;
DR	InterPro: IPR000494; EGFR_L_domain.
DR	InterPro: IPR002174; Furin-1-like.
DR	Pfam: PF00757; Furin-1-like; 1.
DR	Pfam: PF01030; Recep_L_domain; 1.
DR	SMART: SM00261; FU; 2.
QO	SEQUENCE 331 AA; 36489 MW; 45B8EBE683FE7E8 CRC64;

Query Match	32.1%;	Score 734;	DB 4;	Length 331;
Best Local Similarity	44.0%;	Pred. No. 4.7e-59;		
Matches 147;	Conservative 49;	Mismatches 120;	Indels 18;	Gaps 8;

QY	10	GILLALLPFGAA--SPVOGCTDMKLLPASPENHIDMLHHLYGCGVGOVNEIETLPLT	67
	11	GLLESLARGSEYGVSAQVCPOTLNGLSVTDGAENOYOTLKKIYERCEYVWGNLEIYLTGH	70
Db			
QY	68	NASLSFLDIOEVOGYVLLAHNOVROYLPORLRIVRGTOLEEDNYALALVLDNGDPLNNTT	127
	71	NADSLFLDIOEVOGYVLLAHNOVROYLPORLRIVRGTOYDYSKFAFLFVM----	125
Db			
QY	128	PVTGASPGGLREIOLRSITETLLKGGVLIQRRPOLCQYODDTIMKMDIFKNNQOLVLTIDTN	187
Db	126	-----NSSHALRQLRLTQLTETELSSGYLIEKNDKCHMDTIDMDIVARDR-----	178
QY	188	RSRACHPCSPMKCKSRGWGESSEPCGSLTRIVCAGGC-ARCKGPLPLDCCHEOCAGACTG	246
Db	179	NGRSCPPCHEYCKGK-KCWGPGESECCQTLTKTICAPQCGNGCHFGNPQOCCCHDCAGGCSG	237
QY	247	PKHSDCLACLHFNISGICELHCPALVLYNTDIFESMPENEGRYTFGASCTYACRYNLTST	306
Db	238	PQDIDFCACRHFNDSGACVPRCPRLVYNNKLTLPLEENPHRTKTYGGGVCAASPHNFV-V	296
QY	307	DVSGCTLVCPRLHNOEYVAEDSTQCEKSGCKPACR	340
Db	297	DQTSVACRACPPDKMEVD-KNGLKKCECGGGGLCPK	329

RESULT	ID	PRELIMINARY	PRT	AA
99BG66	ID 99BG66			
AC	99BG66;			
DT	01-JUN-2001 (TEMBLrel. 17, Created)			
DT	01-JUN-2001 (TEMBLrel. 17, last sequence update)			
DT	01-OCT-2001 (TEMBLrel. 18, last annotation update)			
OS	Receptor tyrosine kinase ErbB2 (Fragment).			
OC	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Crnata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
NCBI	NCBI_TaxID=9986;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RA	Klonisch T., Wolf P., Hombach-Klonisch S., Vogt S., Kuechenhoff A.,			
RA	Telens F., Fischer B.;			
RT	"ErbB genes and epidermal growth factor- (EGF-) like ligands in the			
RT	peri-implantation rabbit uterus and blastocyst.";			
RL	Submitted (JAN-2001) to the EMBL/Genbank/DBD databases.			
DR	EMBL: AF331378; AAK1471.1; -			
DR	InterPro: IPR002174; Furin-like.			
DR	Pfam: PF00757; Furin-like; 1.			
DR	SMART; SM00261; FU; 2.			
KW	kinase.			
FT	NON_TER	1	1	
FT	NON_TER	149	149	
SQ	SEQUENCE	149 AA;	16240 MW;	7CB3792A54FC9BA CRC64;

Query Match	31.6%	Score 723	DB 6	Length 149
Best Local Similarity	81.2%	Pred. No. 1.8e-58		
Matches 121	Conservative 8	Mismatches 20	Indels 0	Gaps 0

  

OY	158	NPOLCYODTILMKRDIFFHNNQLALDLIDNRSRACHPCSPMKGRCWGESSEDDCOSILNR	217
Db	1	NPQCYOOTILMQEESTRTTSMPSRINASRRARTPCSPCAQAGCWGESSEDDCOSILNR	60
OY	218	TVCAAGCARGCAGPLPTDCCHBQCAAGCTGPKHSDCLACIHNHSGICELHCPALVYNTD	277
Db	61	TTCAGGCGARCGGQLEPTDCCHBQCAAGCTGPKHSDCLACIHNHSGICELHCPALVYNTD	120
OY	278	TFESMPNPEGRTFGASCVTACPNVYLST	306
Db	121	TFESMPNPEGRTFGASCVTACPNVYLST	149

RESULT	11
09YH40	PRELIMINARY; PRT: 1165 AA.
AC 09YH40;	
DT 01-MAY-1999 (TREMBLrel. 10, Created)	
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)	
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE Receptor tyrosine kinase proto-oncogene.	
GN XMRK;	
OS Xiphophorus xiphidium.	
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi	
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei	
OC Acanthomorphi; Acanthopterygii; Perciformes; Atherinomorpha;	
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.	
OX NCBI_Taxid=8086;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN-RIO PURIFICATION;	
RC MEDLINE=98241172; PubMed=9582016;	
RA Dimittijevic N., Winkler C., Welbrock C., Gomez A., Duschl J.,	
RA Altschmid J., Scharlt M.;	
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by	
RT overexpression and mutational alterations.",	
RL Oncogene 16:1681-1690(1998).	

RP SEQUENCE FROM N.A.  
 RC STRAIN-RIO PURIFICATION;  
 RA Schartl M.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U53471; AAD10500.2; ..  
 DR HSP; P11362; 1FGK.  
 DR InterPro: IPR000345; CytC\_heme\_bind.  
 DR InterPro: IPR000494; EGFR\_Ldomain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF00069; kinase; 1.  
 DR Pfam: PF01030; Recp\_Ldomain; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR PRODOM: PD000001; Euk\_kinase; 1.  
 DR SMART: SM00261; FU; 3.  
 DR SMART: SM00219; TYRK; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 DR PROSITE: PSS0107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PSS0011; PROTEIN\_KINASE\_DOM; 2.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR ATP-binding; Kinase; transferase; tyrosine-protein kinase.  
 KW SEQUENCE 1165 AA; 129614 MW; 7f7EE38D8771A74E CRC64;

Query Match	30.3%;	Score 693;	DB 13;	Length 1165;
Best Local Similarity	42.8%;	Pred. No. 1.3e-54;		
Matches 148;	Conservative 46;	Mismatches 134;	Indels 18;	Gaps 9

```

QY 1 MELALACRMGLLLALLPPG -AASr-----OYOTGDMKLRLPASETHLMDLRHLYOCQY 55
Db 4 LELLLEL-----LLLELLSTIGRCSSDPDRKVCYOGTSNOMTM-----LDNHLYMKKKMYSCCNV 56

QY 56 VQGNLELTLYLTNASTLSTLODIOEVOCYVLLAHNOVRPLORLRIARGTOLEFDYNALA 115
Db 57 VLENLEITYYQENDLSTLOSIOEVGVGLYIAMNEVSTIPLYNLRLLRGLNLYEGNTLL 116

QY 116 VLDNGDPLNNTPYVTGASPGGLARELQRLSTLEILLKGVLIQNRPOLCYODTILMKDIFHK 175
Db 117 VMSNYQK-NPSSP--DYVQYVKLQLOLSNLEILLSTIGSVKYSHPNLLCNVETINMMDDIVDK 173

QY 176 NNQALALTLIDTNRBACHPCSPMKKGRMCSESESDQSLTRYCYAGGC-ARCKGPLPTD 234
Db 174 TSNPTMNLIIPIAFEROCOKCDPCGVNSSCAPAGGHOCOKFTKLKACDCNRRCGPGPID 233

```



Db 534 PGPDHMKCAHFIDGPHCVKACPAGV 559

# RESULT 14

Q9BIH9 PRELIMINARY; PRT: 1433 AA.  
 AC Q9BIH9;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative epidermal growth factor receptor (Fragment).  
 GN EGFR.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Anopheles.  
 OX NCBI\_TaxId=7165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SU4;  
 RA Lycett G.J.;  
 RT "Cloning, expression and localisation of the Anopheles gambiae  
 epidermal growth factor receptor";  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AJ301653; CAC35008.1; -;  
 DR HSP; P11362; 1FGK.  
 DR InterPro: IPR000345; Cytochrome\_bind.  
 DR InterPro: IPR000494; EGFR\_Ldomain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF00069; kinase; 1.  
 DR Pfam: PF01030; Recep\_Ldomain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00261; FU; 7.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TyKc; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_4.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW Receptor.  
 FT NON\_TER  
 SQ SEQUENCE 1433 AA; 159585 MW; E3DD98967724F07 CRC64;

Query Match 25.0%; Score 571.5; DB 5; Length 1433;  
 Best Local Similarity 37.2%; Pred. No. 2.4e-43;  
 Matches 120; Conservative 47; Mismatches 125; Indels 31; Gaps 7;

QY 26 CIGTGMKLLPASPETHMLRLHYGCGVGNLLELYPTMASISFLDIQIOEYGYVL 85  
 DB 1 CIGTGMKRVSPANREHYNLDRYNTCTYVDGNETITWIONTTDINPTQHREVGYVL 60  
 QY 86 IAHNOYQVPLORLIRVSTOLF-----EDNYALAVLDNGDPLNNTT PYTGASPGGLREL 140  
 DB 61 ISLYDLPQVILRLQIIRRTTFKLKMKWEAYGLFV-----SFSIMNTL 104  
 QY 141 QLRSLTEILKGVLIQRNPOLCYODITLKKDI-FHKNNOLATLIDTNSRACHPCSPMC 199  
 DB 105 ELPALMDILDIGSGVFNNYLMHKMSINMEELLAPQTSMQTFFNPSPERVCPCHPSC 164  
 QY 200 KSRKMGSESDCCQSLTRVCAAGCA--RKGGLPDPDCCHEGCAACCTGPKHSDCLACH 257  
 DB 165 EVG-CWGEAGHCQRPKSLNCSPQSGRCFGKPRECCCHLCAGGCTGPTQSDCLACKN 223  
 QY 258 FNNSGICEIHCALATLYNDFESMNPGRYTFGASCYTACPYNLTSDVSGCTLVCP 317  
 DB 224 FYDDGVCRCQECRPMQIYNFTNTFMENPNPDGSKATYATCYARKCP-EHLIDNGACVAKCKR 282  
 QY 318 HNOEYTAEDGTORCEKSKPCAR 340

Db 283 GRMPQNSE-----CVPCKGVCPR 300

# RESULT 15

Q9W6F6 PRELIMINARY; PRT: 1137 AA.  
 AC Q9W6F6;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Receptor tyrosine kinase (Fragment).  
 GN ERBB4.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxId=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HINDBRAIN;  
 RX MEDLINE=99263203; PubMed=10328884;  
 RA Dixon M., Lumsden A.;  
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in  
 embryonic chick hindbrain";  
 RL Mol. Cell. Neurosci. 13:237-258(1999).  
 DR EMBL; AF121963; AAD31764.1; -;  
 DR HSP; P11362; 1FGK.  
 DR InterPro: IPR000494; EGFR\_Ldomain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001368; TNFR\_c6.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF00069; kinase; 1.  
 DR Pfam: PF01030; Recep\_Ldomain; 1.  
 DR Pfam: PF02757; YLP; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TyKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00652; TNFR\_NGFR\_1; UNKNOWN\_1.  
 KW kinase; Tyrosine-protein kinase.  
 FT NON\_TER  
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 20.3%; Score 463.5; DB 13; Length 1137;  
 Best Local Similarity 44.8%; Pred. No. 1.4e-33;  
 Matches 86; Conservative 24; Mismatches 73; Indels 9; Gaps 5;

QY 161 LCYODITLKKDIFHKNNOLATLIDTNSRACHPCSPMKGSRKMGSESDCCQSLTRIVC 220  
 DB 3 LCFADITHQDIVIRNFWASNFTLVPTNGSSCGGRCHKSCGTG-RCMGPTNHCQTLLTKVC 61  
 QY 221 AGGC-ARCGPLPPTCCCHQCAAGCTGPKHSDCLACHHHHSIGIELHCPALVTYNTDF 279  
 DB 62 ABOCDGRCYGPVSDCHRECAAGCGSPKDYDCFACMNFNDGACVTCQPGFFVNPPTF 121  
 QY 280 ESMNPGRYTFGASCVTACPYNLTSDVSGCTLVCPLNHOEYTAEDGTORCEKSKPCA 339  
 DB 122 QLEHNINAKYTTGATCYVKKCPHNFV-VDSSSCVACRCPSSKMEV-EEHGKMKRPTDIDCP 179  
 QY 340 R-----GTHSL 346  
 DB 180 KACDGIIGTSLV 191

Search completed: March 4, 2003, 12:36:02  
 Job time: 50.2751 secs

1



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:32:36 ; Search time 22.7169 seconds  
(without alignments)  
542.689 million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287

Sequence: 1 METALACRWGLLALPPGA.....VGRGPDPAHVAVNLRYEG 419

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfillsl.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2287	100.0	419	4	US-09-630-155-2	Sequence 2, Appl1
2	1878	82.1	782	2	US-09-146-283-4	Sequence 4, Appl1
3	1878	82.1	782	3	US-08-579-823A-4	Sequence 4, Appl1
4	1878	82.1	782	4	US-09-344-195-4	Sequence 4, Appl1
5	1878	82.1	1255	1	US-08-467-083-68	Sequence 68, Appl1
6	1878	82.1	1255	1	US-08-414-417B-68	Sequence 68, Appl1
7	1878	82.1	1255	2	US-08-484-438-8	Sequence 68, Appl1
8	1878	82.1	1255	2	US-08-486-348A-68	Sequence 68, Appl1
9	1878	82.1	1255	2	US-08-625-101-2	Sequence 2, Appl1
10	1878	82.1	1255	2	US-08-468-545B-68	Sequence 68, Appl1
11	1878	82.1	1255	3	US-08-356-786-2	Sequence 2, Appl1
12	1878	82.1	1255	3	US-08-466-680B-68	Sequence 68, Appl1
13	1769	77.4	624	3	US-08-422-108-1	Sequence 1, Appl1
14	1769	77.4	624	4	US-08-422-734-1	Sequence 1, Appl1
15	793	34.7	644	1	US-08-336-708A-9	Sequence 1, Appl1
16	793	34.7	1210	2	US-08-484-438-7	Sequence 7, Appl1
17	793	34.7	1210	2	US-08-475-035-4	Sequence 4, Appl1
18	775	33.9	911	2	US-08-484-438-10	Sequence 10, Appl1
19	775	33.9	1058	2	US-08-484-438-4	Sequence 4, Appl1
20	775	33.9	1308	2	US-08-484-438-2	Sequence 2, Appl1
21	773	33.8	478	4	US-09-570-454-2	Sequence 2, Appl1
22	735.5	32.2	1342	1	US-07-978-895-4	Sequence 4, Appl1
23	735.5	32.2	1342	2	US-08-484-438-9	Sequence 9, Appl1
24	735.5	32.2	1342	2	US-08-473-119-4	Sequence 4, Appl1
25	735.5	32.2	1342	2	US-08-475-352-4	Sequence 4, Appl1
26	734	32.1	1343	6	5183884-4	Patent No. 5183884
27	493	21.6	97	1	US-08-421-356-3	Sequence 3, Appl1

28	493	21.6	97	4	US-09-046-783-3	Sequence 3, Appl1
29	418	18.3	79	4	US-09-630-155-1	Sequence 1, Appl1
30	264.5	11.6	1382	2	US-08-737-715-2	Sequence 2, Appl1
31	264.5	11.6	1382	4	US-09-457-040B-7	Sequence 7, Appl1
32	257.5	11.3	1367	2	US-08-746-559A-4	Sequence 4, Appl1
33	257.5	11.3	1367	2	US-08-249-687C-2	Sequence 2, Appl1
34	257.5	11.3	1367	2	US-08-625-819-2	Sequence 2, Appl1
35	257.5	11.3	1367	2	US-08-746-559A-2	Sequence 2, Appl1
36	257.5	11.3	1367	4	US-08-864-641B-18	Sequence 18, Appl1
37	241.5	10.6	486	3	US-08-746-559A-5	Sequence 5, Appl1
38	210.5	9.2	383	4	US-08-857-076-105	Sequence 105, App
39	203	8.9	1724	4	US-08-857-076-12	Sequence 12, Appl1
40	196	8.6	366	4	US-08-857-076-103	Sequence 103, App
41	184.5	8.1	370	4	US-08-857-076-104	Sequence 104, App
42	147.5	6.4	381	4	US-08-857-076-106	Sequence 106, App
43	142	6.2	1940	2	US-08-644-271-30	Sequence 30, Appl1
44	142	6.2	1940	2	US-09-077-955-34	Sequence 34, Appl1
45	131.5	5.7	799	2	US-08-525-940-23	Sequence 23, Appl1

#### ALIGNMENTS

```
RESULT 1
US-09-630-155-2
; Sequence 2, Application US/09630155
; Patent No. 6414130
GENERAL INFORMATION:
APPLICANT: Donerty, Joni Kristin and Gail M. Clinton
TITLE OF INVENTION: HER-2 BINDING ANTAGONISTS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSER: DAVIS WRIGHT TREMAINE LLP
STREET: 1501 Fourth Avenue, 2600 Century Square
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: PC compatible
OPERATING SYSTEM: Windows95
SOFTWARE: word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/630,155
FILING DATE: 16-Jan-2001
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Davison, Barry L.
REGISTRATION NUMBER: 47,309
REFERENCE/DOCKET NUMBER: 49321-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206 628-7621
TELEFAX: 206 628-7699
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 419
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: polypeptide
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-630-155-2
Query Match 100.0%; Score 2287; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 2.8e-192;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 METALACRWGLLALPPGAASVQCTGTGDMKRLRLASPTTHIDMLRHLYOGGVQGNL 60
Db 1 METALACRWGLLALPPGAASVQCTGTGDMKRLRLASPTTHIDMLRHLYOGGVQGNL 60
OY 61 ELTYLPTNASTLFTQDIQEVGVLLIAHNQVROVPIQRLRTVKGTOLEFDNVALAVLDNG 120
```

```
|||||
Db 61 ELTYLPTNASTSLFQDIOEQVGYVLLAHNOVROVPLQRLRIVRGTOLEFEDNALVALVDNG 120
QY 121 DPLNNTPTVYGASPGGLRELOQLRSLEITLKGVLIQNPQLCYQDTILMKDIFHKNNOLA 180
Db 121 DPLNNTPTVYGASPGGLRELOQLRSLEITLKGVLIQNPQLCYQDTILMKDIFHKNNOLA 180
QY 181 LTLIDTNRSPACHPCSPMCKGSRCKWGESSEDDCSLTRTYCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSPACHPCSPMCKGSRCKWGESSEDDCSLTRTYCAGGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDFESMPNPEGRTYFGASCYTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDFESMPNPEGRTYFGASCYTACP 300
QY 301 YNLTSDVGSCTLVCPHLNDEVTAEADGTQRCCKSKPCANGTSLPRPAVPPVPLRMOP 360
Db 301 YNLTSDVGSCTLVCPHLNDEVTAEADGTQRCCKSKPCANGTSLPRPAVPPVPLRMOP 360
QY 361 GPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPISPVSGRGPDPDAHVAVNLSTRYG 419
Db 361 GPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPISPVSGRGPDPDAHVAVNLSTRYG 419
```

## RESULT 2

```
US-09-146-283-4
; Sequence 4, Application US/09146283
; Patent No. 5976546
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Compositions
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,283
; FILING DATE: 03-SEPT-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
US-09-146-283-4
```

```
Query Match 82.1%; Score 1878; DB 2: Length 782;
Best Local Similarity 83.0%; Pred. No. 4e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
```

```
QY 1 METALACRMGLLLALLPFGASTQVCTGTDMLRLPASPETHLDMRLHLYXGCGQVQGNL 60
Db 1 METALACRMGLLLALLPFGASTQVCTGTDMLRLPASPETHLDMRLHLYXGCGQVQGNL 60
QY 61 ELTYLPTNASTSLFQDIOEQVGYVLLAHNOVROVPLQRLRIVRGTOLEFEDNALVALVDNG 120
Db 61 ELTYLPTNASTSLFQDIOEQVGYVLLAHNOVROVPLQRLRIVRGTOLEFEDNALVALVDNG 120
QY 121 DPLNNTPTVYGASPGGLRELOQLRSLEITLKGVLIQNPQLCYQDTILMKDIFHKNNOLA 180
Db 121 DPLNNTPTVYGASPGGLRELOQLRSLEITLKGVLIQNPQLCYQDTILMKDIFHKNNOLA 180
QY 181 LTLIDTNRSPACHPCSPMCKGSRCKWGESSEDDCSLTRTYCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSPACHPCSPMCKGSRCKWGESSEDDCSLTRTYCAGGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDFESMPNPEGRTYFGASCYTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDFESMPNPEGRTYFGASCYTACP 300
QY 301 YNLTSDVGSCTLVCPHLNDEVTAEADGTQRCCKSKPCAR-----GTHSLPRPAVPPV 355
Db 301 YNLTSDVGSCTLVCPHLNDEVTAEADGTQRCCKSKPCARVCGYGLMEHLREVRAVTSAN 360
QY 361 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVPI-----SPVSGRGP 405
Db 361 IOEFACGCKRTFGSLATLPREFDGPASNT---APLOPQLOVETLEITGLYXISAWPD 417
QY 406 --PDAAVAVNLSRYEG 419
Db 418 SLDPDSVFQNLQVIRG 433
```

## RESULT 3

```
US-08-579-823A-4
; Sequence 4, Application US/08579823A
; Patent No. 6080409
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Composition and Method
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/579,823A
; FILING DATE: 03-DEC-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
```



```

; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
US-08-579-823A-4

Query Match
Best local Similarity 82.1%; Score 1878; DB 3; Length 782;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METALCRWGGLLALLPRAASTOVCTGDMKRLRPASPEFHLDMLRHLVYGGQVVOGML 60
DB 1 METALCRWGGLLALLPRAASTOVCTGDMKRLRPASPEFHLDMLRHLVYGGQVVOGML 60
QY 61 ELTYLPTNLSLFLDIOIEVOGVYLLAHNOVROVPLQRLRIYVGTOLFEEDNVALAVLDNG 120
DB 61 ELTYLPTNLSLFLDIOIEVOGVYLLAHNOVROVPLQRLRIYVGTOLFEEDNVALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
QY 181 LTLIDNRSRACHPCSPMKGSRKMGESSEDCOSLTRIVCAGGCARCKGPLPTDCCHEOC 240
DB 181 LTLIDNRSRACHPCSPMKGSRKMGESSEDCOSLTRIVCAGGCARCKGPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMNPBGRTTFGASCYTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMNPBGRTTFGASCYTACP 300
QY 301 YNLSLTVGSCITLVCPLHNOEYTAEDGTORCEKSRPCAR----GTHSLRPRAAVPV 355
DB 301 YNLSLTVGSCITLVCPLHNOEYTAEDGTORCEKSRPCAR----GTHSLRPRAAVPV 355
QY 356 LRMQPG--PAHPLYSLFRPSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
DB 356 LRMQPG--PAHPLYSLFRPSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
QY 406 --PDAHVAVNLSRYEG 419
DB 406 --PDAHVAVNLSRYEG 419
QY 418 SLPLDSVFQNLQVIRG 433
DB 418 SLPLDSVFQNLQVIRG 433

RESULT 4
US-09-344-195-4
; Sequence 4, Application US/09344195
; Patent No. 6210662
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Compositions
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/344,195
; FILING DATE: 24-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,283
; FILING DATE: 03-SEPT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
```

```

; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-344-195-4

Query Match
Best local Similarity 82.1%; Score 1878; DB 4; Length 782;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METALCRWGGLLALLPRAASTOVCTGDMKRLRPASPEFHLDMLRHLVYGGQVVOGML 60
DB 1 METALCRWGGLLALLPRAASTOVCTGDMKRLRPASPEFHLDMLRHLVYGGQVVOGML 60
QY 61 ELTYLPTNLSLFLDIOIEVOGVYLLAHNOVROVPLQRLRIYVGTOLFEEDNVALAVLDNG 120
DB 61 ELTYLPTNLSLFLDIOIEVOGVYLLAHNOVROVPLQRLRIYVGTOLFEEDNVALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
QY 181 LTLIDNRSRACHPCSPMKGSRKMGESSEDCOSLTRIVCAGGCARCKGPLPTDCCHEOC 240
DB 181 LTLIDNRSRACHPCSPMKGSRKMGESSEDCOSLTRIVCAGGCARCKGPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMNPBGRTTFGASCYTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMNPBGRTTFGASCYTACP 300
QY 301 YNLSLTVGSCITLVCPLHNOEYTAEDGTORCEKSRPCAR----GTHSLRPRAAVPV 355
DB 301 YNLSLTVGSCITLVCPLHNOEYTAEDGTORCEKSRPCAR----GTHSLRPRAAVPV 355
QY 356 LRMQPG--PAHPLYSLFRPSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
DB 356 LRMQPG--PAHPLYSLFRPSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
QY 406 --PDAHVAVNLSRYEG 419
DB 406 --PDAHVAVNLSRYEG 419
QY 418 SLPLDSVFQNLQVIRG 433
DB 418 SLPLDSVFQNLQVIRG 433

RESULT 5
US-08-467-083-68
; Sequence 68, Application US/08467083
; Patent No. 5726023
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
```



Patent No. 5811098 5780031  
GENERAL INFORMATION:  
APPLICANT: Plowman, Gregory D.  
APPLICANT: Culouscou, Jean-Michel  
APPLICANT: Shoyab, Mohammed  
APPLICANT: Slegail, Clay B.  
APPLICANT: Hellstr m, Ingegerd  
APPLICANT: Hellstr m, Karl E.  
TITLE OF INVENTION: HERA HUMAN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,438  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/323,442  
FILING DATE: 14-OCT-1994  
APPLICATION NUMBER: US 08/150,704  
FILING DATE: 10-NOV-1993  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/981,165  
FILING DATE: 24-NOV-1992  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 5624-230  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-484-438-8

Query Match 82.1%; Score 1878; DB 2; Length 1355;  
Best Local Similarity 83.0%; Pred. No. 7.5e-156;  
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAAICRMGLALLALPPGAASVQVCTGTDMLRLPASPTHTDMLRHLYOGCOVVOGNL 60  
DB 1 MELAAICRMGLALLALPPGAASVQVCTGTDMLRLPASPTHTDMLRHLYOGCOVVOGNL 60  
QY 61 ELTYLPTNASTSLFDIOEVOGVVLAHNOVROVPLQRLRIVGTOLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASTSLFDIOEVOGVVLAHNOVROVPLQRLRIVGTOLFEDNYALAVLDNG 120  
QY 121 DPLNNTPTVVGASPGGLRELQRLSLEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVVGASPGGLRELQRLSLEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMGKSGRCWSSSDCOSLFTTVAGGACARCKGLPDDCCHEOC 240  
DB 181 LTLIDNRSRACHPCSPMGKSGRCWSSSDCOSLFTTVAGGACARCKGLPDDCCHEOC 240

QY 241 AAGCTGPKHSKDLACLAFHNHSGICELHCPALVYNTDFFESMPNDEGRYTFGASCYTACP 300  
DB 241 AAGCTGPKHSKDLACLAFHNHSGICELHCPALVYNTDFFESMPNDEGRYTFGASCYTACP 300  
QY 301 YNYLSTDVSGCTLVCPRLHNOEVTAEADGTQRCCKSPCAR----GTHSLPRAAVPVP 355  
DB 301 YNYLSTDVSGCTLVCPRLHNOEVTAEADGTQRCCKSPCARVCGLGMEHREVAATVSAN 360  
QY 356 LRMPG--PAHPVLSELRPMDLVSAFSLPLAPLSTSPV-----SPVSVGRGPD 405  
DB 361 IQEFAGCKKIFGSLAFIPESFDGPASNT--APLOEQLQVETLEIRGYLIISAMPD 417  
QY 406 --PDAAHVAVNLRYEG 419  
DB 418 SLDPDSLVSFQNLQVIRG 433

RESULT 8  
US-08-486-348A-68  
Sequence 68, Application US/08486348A  
Patent No. 584538  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Disis, Mary L.  
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,348A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-486-348A-68

Query Match 82.1%; Score 1878; DB 2; Length 1355;  
Best Local Similarity 83.0%; Pred. No. 7.5e-156;  
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAAICRMGLALLALPPGAASVQVCTGTDMLRLPASPTHTDMLRHLYOGCOVVOGNL 60  
DB 1 MELAAICRMGLALLALPPGAASVQVCTGTDMLRLPASPTHTDMLRHLYOGCOVVOGNL 60  
QY 61 ELTYLPTNASTSLFDIOEVOGVVLAHNOVROVPLQRLRIVGTOLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASTSLFDIOEVOGVVLAHNOVROVPLQRLRIVGTOLFEDNYALAVLDNG 120  
QY 121 DPLNNTPTVVGASPGGLRELQRLSLEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVVGASPGGLRELQRLSLEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180

Oy	181	LTTIDTNSRACHCSPKCKSRGWSSEEDCOSLTFTVAGGCARCKGPLPTDCCHEOC	240
Db	181	LTLIDTNSRACHCSPKCKSRGWSSEEDCOSLTTVCAGGCARKGPLPTDCCHEOC	240
Oy	241	AAGCTGPKHSDCLACLFHNHSIGICELHCALPALTNTDTFESMPNDEGRYTFGASCVTACP	3000
Db	241	AAGCTGPKHSCLCLACLFHNHSIGICELHCALPALTNTDTFESMPNDEGRYTFGASCVTACP	3000
Oy	301	YNVLSTDVGSCTLVCPHLNOEVTAEEDGOREKCSKFCAR-----GNHSLTPRAANVPV	3555
Db	301	YNVLSTDVGSCTLVCPHLNÖVTAEDGTÖREKCSKFCARCVGLGMHLEVRAYTSAN	3600
Oy	356	LRMOPG--PAHPVLSFLRPNDLVSAFYSLPLDALPTSVPI-----SPVSVGRRPD	4050
Db	361	IOEFAGCKRIKGSLAFLPESEFDGPASMT---APLOEQVQVFETLEITGVLYISAMPD	4170
Oy	406	--PDAAHVAVNLRYEG	419
Db	418	SLPDLSTVFQNLOVTRG	433

RESULT 9  
US-08-625-101-2  
; Sequence 2, Application US/08625101

APPLICANT: Cheever, Martin A.  
 APPLICANT: Disis, Mary L.  
 TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE  
 TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION  
 TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu  
 TITLE OF INVENTION: ONCOGENE IS ASSOCIATED  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SEED and BERRY LLP  
 STREET: 6300 Columbia Avenue, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104-7092  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/625,101  
 FILING DATE: 01-APR-1996  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sharkey, Richard G.  
 REGISTRATION NUMBER: 32,629  
 REFERENCE/DOCKET NUMBER: 920010.448C7  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1255 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 OS-08-625-101-2

Query Match	82.1%;	Score 1878;	DB 2;	Length 1255;
Best Local Similarity	83.0%;	Pred. No. 7.5e-156;		
Matches 362; Conservative	9;	Mismatches 45;	Indels 20;	Gaps 5

Qy 1 MELAALCRWGLLALPPGAASVCTGTDMKLRLEASPETHDMLRHHYQGCYVQGNL 60  
|||||  
Db 1 MELAALCRWGLLALPPGAASVCTGTDMKLRLEASPETHDMLRHHYQGCYVQGNL 60

Oy	61	ELTLYPTNASLSFLODIOEOVGVLIAHNOVROVLORLRIVRGTOLFENNYALAVLDNG	120
Dd	61	ELTLYPTNASLSFLODIOEOVGVLIAHNOVROVLORLRIVRGTOLFEDNYALAVLDNG	120
Oy	121	DLPLNTPVTASPGSLRELOLRLTEILKGVLIQBNPOLCYODTILMMDIFHKNNOLA	180
Dd	121	DLPLNTPVTASPGSLRELOLRLTEILKGVLIQBNPOLCYODTILMMDIFHKNNOLA	180
Oy	181	LTLIDTNSRACHPCSPCKGSRGWSSESDCOSLRTTVACAGGCARCKGFLPTDCCHEOC	240
Dd	181	LTLIDTNSRACHPCSPCKGSRGWSSESDCOSLRTTVACAGGCARCKGFLPTDCCHEOC	240
Oy	241	AAGCGPRHSOCLCLPHNHSGICELHCPALVYNTDTFESMPMBEGRYTFGASCYACP	300
Dd	241	AAGCGPRHSOCLCLPHNHSGICELHCPALVYNTDTFESMPMBEGRYTFGASCYACP	300
Oy	301	YNYLSTDVSGSTLVCPLHNOEVTAEDETORCEKCSKPCAR-----GTHSLRPPAAVPV	355
Dd	301	YNYLSTDVSGSTLVCPLHNOEVTAEDETQREKCSKCARVCYGLGMHRLREVRAYISAN	360
Oy	356	LRMOPG--PAHPVLSELRPSMDVLSAFYSLPLADLSPYSVI-----SPYSVGRGP	405
Dd	361	IOEFAGCKIRIGSLAFLEPSESDGPASNT--APLOEOLOVETTELTEIRIGLYIASMP	417
Oy	406	--PDAHVAVNLSTREYG	419
Dd	418	SLPDLSTVFQNLQVIRG	433

RESULT 10  
US-08-468-545B-68  
; Sequence 68, Application US/08468545B  
; Patent No. 5876712

1 APPLICANT: Cheever, Martin A.  
2 APPLICANT: Disis, Mary L.  
3 TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
4 TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
5 TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED  
6 NUMBER OF SEQUENCES: 69  
7 CORRESPONDENCE ADDRESS:  
8 ADDRESSEE: Seed and Berry LLP  
9 STREET: 6300 Columblia Center, 701 Fifth Avenue  
10 CITY: Seattle  
11 STATE: Washington  
12 COUNTRY: US  
13 ZIP: 98104-7092  
14 COMPUTER READABLE FORM:  
15 MEDIUM TYPE: Floppy disk  
16 COMPUTER: IBM PC compatible  
17 OPERATING SYSTEM: PC-DOS/MS-DOS  
18 SOFTWARE: PatentIn Release #1.0, Version #1.25  
19 CURRENT APPLICATION DATA:  
20 APPLICATION NUMBER: US/08/468,545B  
21 FILING DATE: 06-JUN-1995  
22 CLASSIFICATION: 424  
23 ATTORNEY/AGENT INFORMATION:  
24 NAME: Sharkey, Richard G.  
25 REGISTRATION NUMBER: 32,629  
26 REFERENCE/DOCKET NUMBER: 920010.448C5  
27 TELECOMMUNICATION INFORMATION:  
28 TELEPHONE: (206) 622-4900  
29 TELEFAX: (206) 682-6031  
30 INFORMATION FOR SEQ ID NO: 68:  
31 SEQUENCE CHARACTERISTICS:  
32 LENGTH: 1255 amino acids  
33 TYPE: amino acid  
34 TOPOLOGY: linear  
35  
36 US-08-468-545B-68

Query Match	82.1%;	Score 1878;	DB 2;	Length 1255;
Best Local Similarity	83.0%;	Pred. No. 7.5e+156;		
Matches 362; Conservative	9;	Mismatches 45;	Indels 20;	Gaps 5

```

OY 1 MELALCRMGILLALPPGAASOVCTGDMKRLPASPETHLMDRLHYOGCOVVOGNI 60
    |||
DB 1 MELALCRMGILLALPPGAASOVCTGDMKRLPASPETHLMDRLHYOGCOVVOGNI 60
OY 61 ELTYLPTNALSFLQDIOEVQGYVLLAHNOVROPLORLRIYRGTLQFEDNYALAVLDNG 120
    |||
DB 61 ELTYLPTNALSFLQDIOEVQGYVLLAHNOVROPLORLRIYRGTLQFEDNYALAVLDNG 120
OY 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
    |||
DB 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
OY 181 LFLIDNRSRACHPCSPMKSGSRCKGSESDCOSLRTVCAGGCARCKPLPTDCCHEOC 240
    |||
DB 181 LFLIDNRSRACHPCSPMKSGSRCKGSESDCOSLRTVCAGGCARCKPLPTDCCHEOC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
    |||
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
OY 301 YNLTSDVGSCTLYCPLAHNOEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVP 355
    |||
DB 301 YNLTSDVGSCTLYCPLAHNOEYTAEDGTORCEKSKPCARCYGLQMEHLREYRAVTSAN 360
OY 356 LRMQPG--PAHPVLSFLRPSMDLVSFAFSLPLAPLSPTSVPI-----SPYVGRGPD 405
    |||
DB 356 LRMQPG--PAHPVLSFLRPSMDLVSFAFSLPLAPLSPTSVPI-----SPYVGRGPD 405
OY 406 --PDAAHVAVNLSRYEG 419
    |||
DB 406 --PDAAHVAVNLSRYEG 419
    |||
DB 418 SLPLDSVFQNLQVIRG 433

```

```

RESULT 11
US-08-356-786-2
; Sequence 2, Application US/08356786
; Patent No. 5877305
; GENERAL INFORMATION:
; APPLICANT: Huston, James S.
; APPLICANT: Oppertmann, Hermann
; APPLICANT: Houston, L. L.
; APPLICANT: Ring, David B.
; TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer
; TITLE OF INVENTION: Marker
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault
; STREET: Exchange Place, 53 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,786
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/831,967
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CR-053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-356-786-2

```

```

Query Match      82.1%; Score 1878; DB 2; Length 1255;
Best Local Similarity 83.0%; Pred. No. 7,5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```

```

OY 1 MELALCRMGILLALPPGAASOVCTGDMKRLPASPETHLMDRLHYOGCOVVOGNI 60
    |||
DB 1 MELALCRMGILLALPPGAASOVCTGDMKRLPASPETHLMDRLHYOGCOVVOGNI 60
OY 61 ELTYLPTNALSFLQDIOEVQGYVLLAHNOVROPLORLRIYRGTLQFEDNYALAVLDNG 120
    |||
DB 61 ELTYLPTNALSFLQDIOEVQGYVLLAHNOVROPLORLRIYRGTLQFEDNYALAVLDNG 120
OY 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
    |||
DB 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
OY 181 LFLIDNRSRACHPCSPMKSGSRCKGSESDCOSLRTVCAGGCARCKPLPTDCCHEOC 240
    |||
DB 181 LFLIDNRSRACHPCSPMKSGSRCKGSESDCOSLRTVCAGGCARCKPLPTDCCHEOC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
    |||
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
OY 301 YNLTSDVGSCTLYCPLAHNOEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVP 355
    |||
DB 301 YNLTSDVGSCTLYCPLAHNOEYTAEDGTORCEKSKPCARCYGLQMEHLREYRAVTSAN 360
OY 356 LRMQPG--PAHPVLSFLRPSMDLVSFAFSLPLAPLSPTSVPI-----SPYVGRGPD 405
    |||
DB 356 LRMQPG--PAHPVLSFLRPSMDLVSFAFSLPLAPLSPTSVPI-----SPYVGRGPD 405
OY 406 --PDAAHVAVNLSRYEG 419
    |||
DB 406 --PDAAHVAVNLSRYEG 419
    |||
DB 418 SLPLDSVFQNLQVIRG 433

```

```

RESULT 12
US-08-466-680B-68
; Sequence 68, Application US/08466680B
; Patent No. 6075122
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,680B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.

```

REGISTRATION NUMBER: 32,629  
 REFERENCE/DOCKET NUMBER: 920010.448C4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 68:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1255 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-466-680B-68

Query Match 82.1%; Score 1878; DB 3; Length 1255;  
 Best local Similarity 83.0%; Pred. No. 7.5e-156;  
 Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METALACRMGLLALLPGASSTQVCTGDMKRLRSPETHLDMRLHLYOGCQVGNL 60  
 DB 1 METALACRMGLLALLPGASSTQVCTGDMKRLRSPETHLDMRLHLYOGCQVGNL 60  
 QY 61 ELTYLPTNASLSFLQDIOEVGYVLLAHNOVRVPLQRLRIVRGTOLEFEDNVALAVLDNG 120  
 DB 61 ELTYLPTNASLSFLQDIOEVGYVLLAHNOVRVPLQRLRIVRGTOLEFEDNVALAVLDNG 120  
 QY 121 DPLNNTPTVTGASPGGLRELQLRSLEFIKGVLIQRPOLCYQDTILMKDIFHKNNQLA 180  
 DB 121 DPLNNTPTVTGASPGGLRELQLRSLEFIKGVLIQRPOLCYQDTILMKDIFHKNNQLA 180  
 QY 181 LTLIDNRSRACHPCSPCKSGRCWGSSESDQSLRTVAGGACARCKGLPTDCCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPCKSGRCWGSSESDQSLRTVAGGACARCKGLPTDCCHEQC 240  
 QY 241 AACCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPBGRYTFGASCYTACP 300  
 DB 241 AACCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPBGRYTFGASCYTACP 300  
 QY 301 YNLSTDVSGCTLVCPHLNDEVTAEQTOCEKSKPCAR-----GTHSLPRPAAYVPP 355  
 DB 301 YNLSTDVSGCTLVCPHLNDEVTAEQTOCEKSKPCAR-----GTHSLPRPAAYVPP 355  
 QY 356 LRMQPG--PAHVAVNLISRYEG 419  
 DB 356 LRMQPG--PAHVAVNLISRYEG 419  
 QY 406 --PDAAVAVNLISRYEG 419  
 DB 406 --PDAAVAVNLISRYEG 419  
 QY 418 SLRDLISVFOQLQYIRG 433  
 DB 418 SLRDLISVFOQLQYIRG 433

RESULT 13  
 US-08-422-108-1  
 ; Sequence 1, Application US/08422108  
 ; Patent No. 6015567  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hudziak, Robert M.  
 ; APPLICANT: Shepard, H. Michael  
 ; APPLICANT: Ullrich, Axel  
 ; TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Winpatin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/422,108

FILING DATE: 14-Apr-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/355460  
 FILING DATE: 13-DEC-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/048346  
 FILING DATE: 15-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/354319  
 FILING DATE: 19-MAY-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lee, Wendy M  
 REGISTRATION NUMBER: 00,000  
 REFERENCE/DOCKET NUMBER: 554C2D2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/225-1994  
 TELEFAX: 415/952-9881  
 TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 624 amino acids  
 ; TYPE: Amino Acid  
 ; TOPOLOGY: Linear  
 US-08-422-108-1

Query Match 77.4%; Score 1769; DB 3; Length 624;  
 Best local Similarity 82.2%; Pred. No. 1e-146;  
 Matches 341; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 22 STQVCTGTDKRLRSPETHLDMRLHLYOGCQVGNLELTYPNASLSFLQDIOEVQ 81  
 DB 1 STQVCTGTDKRLRSPETHLDMRLHLYOGCQVGNLELTYPNASLSFLQDIOEVQ 81  
 QY 82 GYVLIHANQVRVPLQRLRIVRGTOLEFEDNVALAVLDNGDPLNNTPTVTGASPGGLRELQ 141  
 DB 61 GYVLIHANQVRVPLQRLRIVRGTOLEFEDNVALAVLDNGDPLNNTPTVTGASPGGLRELQ 120  
 QY 142 LRSLEFIKGVLIQRPOLCYQDTILMKDIFHKNNQLATLIDNRSRACHPCSPCKG 201  
 DB 121 LRSLEFIKGVLIQRPOLCYQDTILMKDIFHKNNQLATLIDNRSRACHPCSPCKG 180  
 QY 202 SRCWGSSESDQSLRTVAGGACARCKGLPTDCCHEQCAAGCTGPKHSDCLACLFHNS 261  
 DB 181 SRCWGSSESDQSLRTVAGGACARCKGLPTDCCHEQCAAGCTGPKHSDCLACLFHNS 240  
 QY 262 GICELHCPALVTYNTDFESMNPBGRYTFGASCYTACPYNLSTDVSGCTLVCPHLNDE 321  
 DB 241 GICELHCPALVTYNTDFESMNPBGRYTFGASCYTACPYNLSTDVSGCTLVCPHLNDE 300  
 QY 322 VTAEQTOCEKSKPCAR-----GTHSLPRPAAYVPLRMQPG--PAHVAVNLISRYEG 374  
 DB 301 VTAEQTOCEKSKPCARCYGLGMEHLREVAAYVSANQIEFGACKTIFGSLAFLEPST 360  
 QY 375 DLVSATYSLPLAPLSPTSVPI-----SPVSVGRPD--PDAAVAVNLISRYEG 419  
 DB 361 DGDPAASNT--APLOPEQLOVFEFTEITEITYLISAMPDLSVFOQLQYIRG 412

RESULT 14  
 US-08-422-734-1  
 ; Sequence 1, Application US/08422734  
 ; Patent No. 6333169  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hudziak, Robert M.  
 ; APPLICANT: Shepard, H. Michael  
 ; APPLICANT: Ullrich, Axel  
 ; TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco

```
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPacIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/422,734
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/422108
FILING DATE: 14-Apr-1995
APPLICATION NUMBER: 08/355460
FILING DATE: 13-Dec-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/048346
FILING DATE: 15-Apr-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/354319
FILING DATE: 19-May-1989
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 554C2D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 624 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-422-734-1

Query Match          77.4%; Score 1769; DB 4; Length 624;
Best Local Similarity 82.2%; Pred. No. 1e-146;
Matches 341; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 22 STGYCTGTDMLRLPASPETHLMDLRHLVYGCQVYVGNLELTYLPTNASLFLDIDIEVQ 81
DB 1 STGYCTGTDMLRLPASPETHLMDLRHLVYGCQVYVGNLELTYLPTNASLFLDIDIEVQ 60
QY 82 GYVLIANOVROYPLQRLIRYRGTOLEFEDNYALAVLNDGDPPLNTPVYTGASPGELRELQ 141
DB 61 GYVLIANOVROYPLQRLIRYRGTOLEFEDNYALAVLNDGDPPLNTPVYTGASPGELRELQ 120
QY 142 LRSITELKGGVLIORNPOLCYODTILMKDIFHKNNOLATLTLIDFNRSRACHPCSPMCKG 201
DB 121 LRSITELKGGVLIORNPOLCYODTILMKDIFHKNNOLATLTLIDFNRSRACHPCSPMCKG 180
QY 202 SRWGSESSDCQSLTRTVACGAGCARCKGPLPTDCHDQCAAGCTGPKHSDCLACILHNHS 261
DB 181 SRWGSESSDCQSLTRTVACGAGCARCKGPLPTDCHDQCAAGCTGPKHSDCLACILHNHS 240
QY 262 GICELHCPALVYNTDTFESMPNDEGRYTTGASCVTACPPYNYLSTDVGSCTLYCPPLHNOE 321
DB 241 GICELHCPALVYNTDTFESMPNDEGRYTTGASCVTACPPYNYLSTDVGSCTLYCPPLHNOE 300
QY 322 VTAEDGTORCEKSKPCAR-----GTHSLRPRAAVVPLRMOPG--PAHPVLSFLRPSW 374
DB 301 VTAEDGTORCEKSKPCARCYGILGMEHLREVRVTSANIOERAGCKKIFGSLAEFLPSF 360
QY 375 DLVSATYSLPLAPLSPTSPVPI-----SPVSVGRGPD--PDAHVAVNLSTRYEG 419
DB 361 DGDPASNT---APLQPEQLQVFTLEETLEYIVISAMPDLSPLDSVQNQLQVIRG 412
```

RESULT 15  
US-08-336-708A-9

```
Sequence 9, Application US/08336708A
Patent No. 5521295
GENERAL INFORMATION:
APPLICANT: Pacifici, Robert E.
APPLICANT: Thomson, Arlen R.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: Hybrid Receptor Molecules
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Denavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,708A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy
REFERENCE/DOCKET NUMBER: A-241A
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 644 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-708A-9

Query Match          34.7%; Score 793; DB 1; Length 644;
Best Local Similarity 45.3%; Pred. No. 3e-61;
Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

QY 11 LLLALPPGAA--STGYCTGTDMLRLPASPETHLMDLRHLVYGCQVYVGNLELTYLPTN 68
DB 14 LLAALCPASRALEEKKYCGQTSNKLTLQGTGFEDHFLSLQRFNCEVVLGSLTEITYVGRN 73
QY 69 ASLSFLDIDIEVGYVLIANOVROYPLQRLIRYRGTOLEFEDNYALAVLNDGDPPLNTP 128
DB 74 YDLSFKTIDIEVAGVYVLIANOVROYPLQRLIRYRGTOLEFEDNYALAVLNDGDPPLNTP 126
QY 129 VTGASPGELRELQRLSITELKGGVLIORNPOLCYODTILMKDIFHKNNOLATLTLIDTNR 188
DB 127 ---ANKGLKELEPHRNIOELIHGAVRNSNPALCNVESIQWRDIVSSDFLSNMSMDFOH 183
QY 189 SRACHPCSPMCKGSRWGSESSDCQSLTRTVACGAGCA-RCKGPLPTDCHDQCAAGCTGP 247
DB 184 LGSQCKDPCSPMCKGSRWGSESSDCQSLTRTVACGAGCA-RCKGPLPTDCHDQCAAGCTGP 243
QY 248 KHSQCLACILHNSGTELEHCPALVYNTDTFESMPNDEGRYTTGASCVTACPPYNYLSTD 307
DB 244 RESQCLVCRKFRDEATKDKCPPLMLYNTPTQMDVBNDEGRYSFGATCVKCPRYNYVTD 303
QY 308 VGSCTLYCPPLHNOEVTAEEDGTORCEKSKPCAR 340
DB 304 HGSCVTRACGADSYEM-EDDGVRRCKKCEGFCRK 335
```

Search completed: March 4, 2003, 12:37:12  
Job time : 26.7169 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:36:06 ; Search time 17.6687 Seconds  
(without alignments)  
1000.035 Million cell updates/sec

Title: US-09-234-208b-2  
Perfect score: 2287  
Sequence: 1 METALCRWGLLALLPFGA.....YGRGPRDAHVAVLSRYEG 419

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_AA: \*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep: \*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep: \*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep: \*  
4: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep: \*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep: \*  
6: /cgn2\_6/ptodata/2/pubpaa/PCT07\_PUBCOMB pep: \*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep: \*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep: \*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep: \*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep: \*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep: \*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep: \*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep: \*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	82.1	645	10	US-09-921-161-1
2	1878	82.1	653	9	US-09-854-356-3
3	1878	82.1	712	9	US-09-854-356-7
4	1878	82.1	919	9	US-09-854-356-6
5	1878	82.1	1255	9	US-09-769-508-2
6	1878	82.1	1255	9	US-09-854-356-1
7	1878	82.1	1255	9	US-09-930-125-2
8	1878	82.1	1255	9	US-09-441-411-6
9	1878	82.1	1255	10	US-09-811-113-9
10	1878	82.1	1255	10	US-09-811-113-3
11	1612.5	70.5	479	10	US-09-821-883-5
12	1610.5	70.4	564	10	US-09-821-883-3
13	1610.5	70.4	697	10	US-09-821-883-4
14	1608.5	70.3	654	9	US-09-854-356-8
15	1608.5	70.3	1256	9	US-09-854-356-2
16	1608.5	70.3	1260	9	US-09-870-759-118
17	1607	70.3	555	10	US-09-821-883-1
18	1607	70.3	690	10	US-09-821-883-2
19	1597.5	69.9	1256	9	US-09-854-356-14

20	1587	69.4	289	10	US-09-821-883-23	Sequence 23, App1
21	987	43.2	191	9	US-09-441-411-9	Sequence 9, App1
22	793	34.7	1210	10	US-09-725-433-2	Sequence 2, App1
23	775	33.9	1308	10	US-09-940-101-2	Sequence 2, App1
24	773	33.8	478	10	US-09-867-521-2	Sequence 2, App1
25	771	33.7	615	10	US-09-940-101-4	Sequence 4, App1
26	257.5	11.3	1367	9	US-09-870-759-120	Sequence 120, App
27	210.5	9.2	383	10	US-09-205-658-105	Sequence 105, App
28	210.5	9.2	383	10	US-09-844-353A-105	Sequence 105, App
29	203	8.9	1724	10	US-09-205-658-12	Sequence 12, App1
30	203	8.9	1724	10	US-09-844-353A-12	Sequence 12, App1
31	196	8.6	366	10	US-09-205-658-103	Sequence 103, App
32	196	8.6	366	10	US-09-844-353A-103	Sequence 103, App
33	184.5	8.1	370	10	US-09-205-658-104	Sequence 104, App
34	184.5	8.1	370	10	US-09-844-353A-104	Sequence 104, App
35	147.5	6.4	381	10	US-09-205-658-106	Sequence 106, App
36	147.5	6.4	381	10	US-09-844-353A-106	Sequence 106, App
37	142	6.2	1940	9	US-10-016-283-34	Sequence 34, App1
38	133.5	5.8	420	9	US-09-796-753-86	Sequence 86, App1
39	130.5	5.7	833	9	US-10-226-296-5	Sequence 5, App1
40	129.5	5.7	830	9	US-09-870-759-140	Sequence 140, App
41	128.5	5.6	420	9	US-09-905-291A-109	Sequence 109, App
42	128.5	5.6	420	9	US-09-902-853-109	Sequence 109, App
43	128.5	5.6	420	9	US-09-907-824-109	Sequence 109, App
44	128.5	5.6	420	9	US-09-907-841-109	Sequence 109, App
45	128.5	5.6	420	9	US-09-904-011-109	Sequence 109, App

#### ALIGNMENTS

RESULT 1  
US-09-921-161-1  
Sequence 1, Application US/09921161  
Patent No. US2002090662A1  
GENERAL INFORMATION:  
APPLICANT: Ralph, Peter  
TITLE OF INVENTION: ANALYTICAL METHOD  
FILE REFERENCE: GENENT.066A  
CURRENT APPLICATION NUMBER: US/09/921,161  
CURRENT FILING DATE: 2001-08-01  
PRIOR APPLICATION NUMBER: 60/225,433  
PRIOR FILING DATE: 2000-08-15  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 645  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-921-161-1

QY	1	METALCRWGLLALLPFGAASOVCTGTMKRLRASPETHIDMLRHLYOGGOVVOGNL	60
DB	1	METALCRWGLLALLPFGAASOVCTGTMKRLRASPETHIDMLRHLYOGGOVVOGNL	60
QY	61	ELTYLPTNASTFLQDIOEVGYVLLAHNOVROVPLQRLRTIVGTOLFEDNYALAVDNG	120
DB	61	ELTYLPTNASTFLQDIOEVGYVLLAHNOVROVPLQRLRTIVGTOLFEDNYALAVDNG	120
QY	121	DPANNTPTVAGSPGGIREQLRLSLTEILKGVLIQRNPOLCYODITLWKDITRHNKQOLA	180
DB	121	DPANNTPTVAGSPGGIREQLRLSLTEILKGVLIQRNPOLCYODITLWKDITRHNKQOLA	180
QY	181	LTLIDNRSRACHPCSPMGCRGWSSESDCOSLRTVAGGCARCKGFLPDDCCEOC	240
DB	181	LTLIDNRSRACHPCSPMGCRGWSSESDCOSLRTVAGGCARCKGFLPDDCCEOC	240
QY	241	AACTGPKHSDCIACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP	300
DB	241	AACTGPKHSDCIACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP	300

```
Db 241 AAGCTGKHSDDLACLFHFNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
QY 301 YNTISTVGSCTTLCPLHNEVTAEDGTORCEKSKPCAR-----GTHSLPRPAAVPV 355
Db 301 YNTISTVGSCTTLCPLHNEVTAEDGTORCEKSKPCARCYGLGMEHLREVAAYTSAN 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
Db 361 IQEFAAGCKKIFGSLAFLEPESFDGDPASNT---APLOPEQLQVETLEITIGYLXISAMPD 417
QY 406 --PDAAVAVNLSRREG 419
Db 418 SLPDLVSFQNLQYIRG 433

RESULT 2
US-09-854-356-3
; Sequence 3, Application US/09854356
; Patent No. US20020177567A1
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Corixa Corporation
; APPLICANT: SmithKline Beecham Biologicals S. A.
; TITLE OF INVENTION: HER-2/neu Fusion Proteins
; FILE REFERENCE: 014058-009810PC
; CURRENT APPLICATION NUMBER: US/09/854,356
; CURRENT FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: US 09/493,480
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: US 60/117,976
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: extracellular domain (ECD) of human HER-2/neu
US-09-854-356-3

Query Match 82.1%; Score 1878; DB 9; Length 653;
Best Local Similarity 83.0%; Pred. No. 1.6e-130;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMKRLRSPSPETHDMLRHLYOGGCOVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMKRLRSPSPETHDMLRHLYOGGCOVQGNL 60
QY 61 ELTYLPTNASTSLFDIOEVOGYVLIHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
Db 61 ELTYLPTNASTSLFDIOEVOGYVLIHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTTPVVGASPGGLREQLRSLEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db 121 DPLNNTTPVVGASPGGLREQLRSLEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
QY 181 LTIIDTNRSRACHPCSPMCKGSRCKWGESSEDCOSLRTTVACAGGACARCKGPLPTDCHEQC 240
Db 181 LTIIDTNRSRACHPCSPMCKGSRCKWGESSEDCOSLRTTVACAGGACARCKGPLPTDCHEQC 240
QY 241 AAGCTGKHSDDLACLFHFNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
Db 241 AAGCTGKHSDDLACLFHFNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
QY 301 YNTISTVGSCTTLCPLHNEVTAEDGTORCEKSKPCAR-----GTHSLPRPAAVPV 355
Db 301 YNTISTVGSCTTLCPLHNEVTAEDGTORCEKSKPCARCYGLGMEHLREVAAYTSAN 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
Db 361 IQEFAAGCKKIFGSLAFLEPESFDGDPASNT---APLOPEQLQVETLEITIGYLXISAMPD 417
```

```
QY 406 --PDAAVAVNLSRREG 419
Db 418 SLPDLVSFQNLQYIRG 433

RESULT 3
US-09-854-356-7
; Sequence 7, Application US/09854356
; Patent No. US20020177567A1
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Corixa Corporation
; APPLICANT: SmithKline Beecham Biologicals S. A.
; TITLE OF INVENTION: HER-2/neu Fusion Proteins
; FILE REFERENCE: 014058-009810PC
; CURRENT APPLICATION NUMBER: US/09/854,356
; CURRENT FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: US 09/493,480
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: US 60/117,976
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
US-09-854-356-7

Query Match 82.1%; Score 1878; DB 9; Length 712;
Best Local Similarity 83.0%; Pred. No. 1.8e-130;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMKRLRSPSPETHDMLRHLYOGGCOVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMKRLRSPSPETHDMLRHLYOGGCOVQGNL 60
QY 61 ELTYLPTNASTSLFDIOEVOGYVLIHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
Db 61 ELTYLPTNASTSLFDIOEVOGYVLIHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTTPVVGASPGGLREQLRSLEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db 121 DPLNNTTPVVGASPGGLREQLRSLEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
QY 181 LTIIDTNRSRACHPCSPMCKGSRCKWGESSEDCOSLRTTVACAGGACARCKGPLPTDCHEQC 240
Db 181 LTIIDTNRSRACHPCSPMCKGSRCKWGESSEDCOSLRTTVACAGGACARCKGPLPTDCHEQC 240
QY 241 AAGCTGKHSDDLACLFHFNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
Db 241 AAGCTGKHSDDLACLFHFNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
QY 301 YNTISTVGSCTTLCPLHNEVTAEDGTORCEKSKPCAR-----GTHSLPRPAAVPV 355
Db 301 YNTISTVGSCTTLCPLHNEVTAEDGTORCEKSKPCARCYGLGMEHLREVAAYTSAN 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
Db 361 IQEFAAGCKKIFGSLAFLEPESFDGDPASNT---APLOPEQLQVETLEITIGYLXISAMPD 417
QY 406 --PDAAVAVNLSRREG 419
Db 418 SLPDLVSFQNLQYIRG 433

RESULT 4
US-09-854-356-6
```

```
; Sequence 6, Application US/09854356
; Patent No. US20020177567A1
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Cheyzen, Dirk
; APPLICANT: Corixa Corporation
; APPLICANT: SmithKline Beecham Biologicals S. A.
; TITLE OF INVENTION: HER-2/neu Fusion Proteins
; FILE REFERENCE: 014058-009810PC
; CURRENT APPLICATION NUMBER: US/09/854,356
; CURRENT FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: US 09/493,480
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: US 60/117,976
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: of ECD and PD of human HER-2/neu
US-09-854-356-6
```

```
Query Match      82.1%; Score 1878; DB 9; Length 919;
Best local Similarity 83.0%; Pred. No. 2,4e-130;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
```

```
QY 1 MELAALCRWGILLALLPRGAASVQVCTGDMKLRPASPEITHMLRHLXGCGVQVGNL 60
DB 1 MELAALCRWGILLALLPRGAASVQVCTGDMKLRPASPEITHMLRHLXGCGVQVGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLLAHNOVQVPLQRLRIYRGTOLEFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLLAHNOVQVPLQRLRIYRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGLREQLRLSTLEILKGVLIQIRNPOLCYODITLTKDIFHKNNOLA 180
DB 121 DPLNNTPTVYGASPGGLREQLRLSTLEILKGVLIQIRNPOLCYODITLTKDIFHKNNOLA 180
QY 121 DPLNNTPTVYGASPGGLREQLRLSTLEILKGVLIQIRNPOLCYODITLTKDIFHKNNOLA 180
DB 121 DPLNNTPTVYGASPGGLREQLRLSTLEILKGVLIQIRNPOLCYODITLTKDIFHKNNOLA 180
QY 181 LFLIDNRSRACHPCSPMKGSRGWGESSDQSLRTVACAGCARGKPLPTDCHEQC 240
DB 181 LFLIDNRSRACHPCSPMKGSRGWGESSDQSLRTVACAGCARGKPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRTYFGASCYTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRTYFGASCYTACP 300
QY 301 YNLSIDVSGCTVCPDLHNOEVAEDGTORCEKSKPCAR-----GTHSLRPAVAVP 355
DB 301 YNLSIDVSGCTVCPDLHNOEVAEDGTORCEKSKPCAR-----GTHSLRPAVAVP 355
QY 301 YNLSIDVSGCTVCPDLHNOEVAEDGTORCEKSKPCARVCGGLMEHLREYRAVTSAN 360
DB 301 YNLSIDVSGCTVCPDLHNOEVAEDGTORCEKSKPCARVCGGLMEHLREYRAVTSAN 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSFAFSLPLAPLSPTSVPI-----SPVSGRGPD 405
DB 356 LRMQPG--PAHPVLSFLRPSMDLVSFAFSLPLAPLSPTSVPI-----SPVSGRGPD 405
QY 361 IOEFAGCKKIFGSLAPLPSFDGDPASNT---APLQPEQLQVEETLEITGYLIISAMPD 417
DB 361 IOEFAGCKKIFGSLAPLPSFDGDPASNT---APLQPEQLQVEETLEITGYLIISAMPD 417
QY 406 --PDAPVAVNLSRYEG 419
DB 406 --PDAPVAVNLSRYEG 419
QY 418 SLPDLSVFQNLQVIRG 433
DB 418 SLPDLSVFQNLQVIRG 433
```

```
RESULT 5
US-09-769-508-2
; Sequence 2, Application US/09769508
; Patent No. US2002015527A1
; GENERAL INFORMATION:
; APPLICANT: STUART, SUSAN G.
; APPLICANT: MONAHAN, JOHN J.
; APPLICANT: LANGTON, BEATRICE CLAUDIA
; APPLICANT: HANCOCK, MIRIAM E.C.
; APPLICANT: CHAO, LORRINE A.
```

```
; APPLICANT: BLUFORD, PETER
; TITLE OF INVENTION: C-ERBB-2 EXTERNAL DOMAIN: GP75
; FILE REFERENCE: BEBIO-111-C1
; CURRENT APPLICATION NUMBER: US/09/769,508
; CURRENT FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-769-508-2
```

```
Query Match      82.1%; Score 1878; DB 9; Length 1255;
Best local Similarity 83.0%; Pred. No. 3,5e-130;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
```

```
QY 1 MELAALCRWGILLALLPRGAASVQVCTGDMKLRPASPEITHMLRHLXGCGVQVGNL 60
DB 1 MELAALCRWGILLALLPRGAASVQVCTGDMKLRPASPEITHMLRHLXGCGVQVGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLLAHNOVQVPLQRLRIYRGTOLEFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLLAHNOVQVPLQRLRIYRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGLREQLRLSTLEILKGVLIQIRNPOLCYODITLTKDIFHKNNOLA 180
DB 121 DPLNNTPTVYGASPGGLREQLRLSTLEILKGVLIQIRNPOLCYODITLTKDIFHKNNOLA 180
QY 121 DPLNNTPTVYGASPGGLREQLRLSTLEILKGVLIQIRNPOLCYODITLTKDIFHKNNOLA 180
DB 121 DPLNNTPTVYGASPGGLREQLRLSTLEILKGVLIQIRNPOLCYODITLTKDIFHKNNOLA 180
QY 181 LFLIDNRSRACHPCSPMKGSRGWGESSDQSLRTVACAGCARGKPLPTDCHEQC 240
DB 181 LFLIDNRSRACHPCSPMKGSRGWGESSDQSLRTVACAGCARGKPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRTYFGASCYTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRTYFGASCYTACP 300
QY 301 YNLSIDVSGCTVCPDLHNOEVAEDGTORCEKSKPCAR-----GTHSLRPAVAVP 355
DB 301 YNLSIDVSGCTVCPDLHNOEVAEDGTORCEKSKPCAR-----GTHSLRPAVAVP 355
QY 301 YNLSIDVSGCTVCPDLHNOEVAEDGTORCEKSKPCARVCGGLMEHLREYRAVTSAN 360
DB 301 YNLSIDVSGCTVCPDLHNOEVAEDGTORCEKSKPCARVCGGLMEHLREYRAVTSAN 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSFAFSLPLAPLSPTSVPI-----SPVSGRGPD 405
DB 356 LRMQPG--PAHPVLSFLRPSMDLVSFAFSLPLAPLSPTSVPI-----SPVSGRGPD 405
QY 361 IOEFAGCKKIFGSLAPLPSFDGDPASNT---APLQPEQLQVEETLEITGYLIISAMPD 417
DB 361 IOEFAGCKKIFGSLAPLPSFDGDPASNT---APLQPEQLQVEETLEITGYLIISAMPD 417
QY 406 --PDAPVAVNLSRYEG 419
DB 406 --PDAPVAVNLSRYEG 419
QY 418 SLPDLSVFQNLQVIRG 433
DB 418 SLPDLSVFQNLQVIRG 433
```

```
RESULT 6
US-09-854-356-1
; Sequence 1, Application US/09854356
; Patent No. US20020177567A1
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Cheyzen, Dirk
; APPLICANT: Corixa Corporation
; APPLICANT: SmithKline Beecham Biologicals S. A.
; TITLE OF INVENTION: HER-2/neu Fusion Proteins
; FILE REFERENCE: 014058-009810PC
; CURRENT APPLICATION NUMBER: US/09/854,356
; CURRENT FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: US 09/493,480
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: US 60/117,976
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```

; FEATURE:
; OTHER INFORMATION: human HER-2/neu protein
; NAME/KEY: DOMAIN
; LOCATION: (1)..(653)
; OTHER INFORMATION: extracellular domain (ECD)
; NAME/KEY: DOMAIN
; LOCATION: (676)..(1255)
; OTHER INFORMATION: intracellular domain (ICD)
; NAME/KEY: DOMAIN
; LOCATION: (990)..(1255)
; OTHER INFORMATION: phosphorylation domain (PD)
; NAME/KEY: DOMAIN
; LOCATION: (990)..(1048)
; OTHER INFORMATION: fragment of the phosphorylation domain, preferred
; OTHER INFORMATION: portion (delta PD)
US-09-854-356-1

Query Match
Best Local Similarity 82.1%; Score 1878; DB 9; Length 1255;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

OY 1 METALCRWGILLALLPPGAASYOCTGDMKRLPASPEHLDMLRHLYOGCQVVGML 60
DB 1 METALCRWGILLALLPPGAASYOCTGDMKRLPASPEHLDMLRHLYOGCQVVGML 60
OY 61 ELTYLPTNASLSFLDIOEVQGVYLAHNOVQVPLQRLRIYRGQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGVYLAHNOVQVPLQRLRIYRGQLFEDNYALAVLDNG 120
OY 121 DPLNNTPTVTGASPGGLRELOLSRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVTGASPGGLRELOLSRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCQSLRTVCAGGACARCKGPLPTDCCHEQC 240
DB 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCQSLRTVCAGGACARCKGPLPTDCCHEQC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRTFGASCVTACP 300
OY 301 YNLSLDVSGSCTLVCPHLNQEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVAVP 355
DB 301 YNLSLDVSGSCTLVCPHLNQEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVAVP 355
OY 356 LRMQPG--PAHPLSLFRSMDLVSAFYSLPLAPLSPTVPI-----SPVSVGRGPD 405
DB 356 LRMQPG--PAHPLSLFRSMDLVSAFYSLPLAPLSPTVPI-----SPVSVGRGPD 405
OY 406 --PDAHVAVNLRSYEG 419
DB 406 --PDAHVAVNLRSYEG 419
OY 418 SLPDLVSFQNLQYIRG 433
DB 418 SLPDLVSFQNLQYIRG 433

RESULT 7
US-09-930-125-2
; Sequence 2, Application US/09930125
; Publication No. US20020193329A1
; GENERAL INFORMATION:
; APPLICANT: Hand-Zimmerman, Susan
; APPLICANT: Cheever, Martin A.
; APPLICANT: Foy, Teresa M.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Kalos, Michael D.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Vedvick, Thomas S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.544
; CURRENT APPLICATION NUMBER: US/09/930,125
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
```

```

; SEQ ID NO 2
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-930-125-2

Query Match
Best Local Similarity 82.1%; Score 1878; DB 9; Length 1255;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

OY 1 METALCRWGILLALLPPGAASYOCTGDMKRLPASPEHLDMLRHLYOGCQVVGML 60
DB 1 METALCRWGILLALLPPGAASYOCTGDMKRLPASPEHLDMLRHLYOGCQVVGML 60
OY 61 ELTYLPTNASLSFLDIOEVQGVYLAHNOVQVPLQRLRIYRGQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGVYLAHNOVQVPLQRLRIYRGQLFEDNYALAVLDNG 120
OY 121 DPLNNTPTVTGASPGGLRELOLSRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVTGASPGGLRELOLSRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCQSLRTVCAGGACARCKGPLPTDCCHEQC 240
DB 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCQSLRTVCAGGACARCKGPLPTDCCHEQC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRTFGASCVTACP 300
OY 301 YNLSLDVSGSCTLVCPHLNQEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVAVP 355
DB 301 YNLSLDVSGSCTLVCPHLNQEYTAEDGTORCEKSKPCAR-----GTHSLPRPAVAVP 355
OY 356 LRMQPG--PAHPLSLFRSMDLVSAFYSLPLAPLSPTVPI-----SPVSVGRGPD 405
DB 356 LRMQPG--PAHPLSLFRSMDLVSAFYSLPLAPLSPTVPI-----SPVSVGRGPD 405
OY 406 --PDAHVAVNLRSYEG 419
DB 406 --PDAHVAVNLRSYEG 419
OY 418 SLPDLVSFQNLQYIRG 433
DB 418 SLPDLVSFQNLQYIRG 433

RESULT 8
US-09-441-411-6
; Sequence 6, Application US/09441411
; Publication No. US20030008342A1
; GENERAL INFORMATION:
; APPLICANT: Scholler, Nathalie B.
; APPLICANT: Disis, Mary L.
; APPLICANT: Hellstrom, Inggerd
; APPLICANT: Hellstrom, Karl Erik
; TITLE OF INVENTION: SURFACE RECEPTOR ANTIGEN VACCINES
; FILE REFERENCE: 730033.409
; CURRENT APPLICATION NUMBER: US/09/441,411
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-441-411-6

Query Match
Best Local Similarity 82.1%; Score 1878; DB 9; Length 1255;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

OY 1 METALCRWGILLALLPPGAASYOCTGDMKRLPASPEHLDMLRHLYOGCQVVGML 60
DB 1 METALCRWGILLALLPPGAASYOCTGDMKRLPASPEHLDMLRHLYOGCQVVGML 60
OY 61 ELTYLPTNASLSFLDIOEVQGVYLAHNOVQVPLQRLRIYRGQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGVYLAHNOVQVPLQRLRIYRGQLFEDNYALAVLDNG 120
```

```
|||||
Db 61 ELTYLPTNASLSFLDIOEVQGYVLAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
|||
Qy 121 DPLNNTPTVYGASPGGLRELOLSLFEILKGVLIORNPOLCYODITLMKDIFFHKNNOA 180
|||
Db 121 DPLNNTPTVYGASPGGLRELOLSLFEILKGVLIORNPOLCYODITLMKDIFFHKNNOA 180
|||
Qy 181 LTLIDTNRSRACHPCSPMKGSRMCWGESSEDCOSLTRTYCAGGCARCKGPLPTDCCHQC 240
|||
Db 181 LTLIDTNRSRACHPCSPMKGSRMCWGESSEDCOSLTRTYCAGGCARCKGPLPTDCCHQC 240
|||
Qy 241 AACCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
|||
Db 241 AACCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
|||
Qy 301 YNLTSDVSGCTLVCPHLHNOEYTAEDGTORCEKCSKPCARVCYGLMEHLREVRAVTSAN 360
|||
Db 301 YNLTSDVSGCTLVCPHLHNOEYTAEDGTORCEKCSKPCARVCYGLMEHLREVRAVTSAN 360
|||
Qy 356 LRMOPG--PAHVPYSLFRPSMDLVSAFYSLPLAPLSPTSVPI-----SPYVGRGPD 405
|||
Db 361 IOEFACCKKIFGSLAFPLPESFDGPASNT---APLOPEOLQVFEITLLETITGLYISAMPD 417
|||
Qy 406 --PDAAVAVNLSRYEG 419
|||
Db 418 SLFDLSVFQNLQYIRG 433
|||
```

## RESULT 9

```
US-09-811-123-9
; Sequence 9, Application US/09811123
; Patent No. US20020001587A1
; GENERAL INFORMATION:
; APPLICANT: Sharon Erickson
; APPLICANT: Ralph Schwalli
; APPLICANT: Mark Slikowski
; TITLE OF INVENTION: METHODS OF TREATMENT USING ANTI-ERBB
; FILE REFERENCE: GENE.073A2
; CURRENT APPLICATION NUMBER: US/09/811,123
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/238,327
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 09/602,530
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-811-123-9
```

```
Query Match      82.1%; Score 1878; DB 10; Length 1255;
Best Local Similarity 83.0%; Pred. No. 3.5e-130;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
```

```
Qy 1 MELAALCRMGILLALPPAASSTOVCTGDMKLRLPASPTHLDMLRHLYOGCOVVOGNTL 60
|||
Db 1 MELAALCRMGILLALPPAASSTOVCTGDMKLRLPASPTHLDMLRHLYOGCOVVOGNTL 60
|||
Qy 61 ELTYLPTNASLSFLDIOEVQGYVLAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
|||
Db 61 ELTYLPTNASLSFLDIOEVQGYVLAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
|||
Qy 121 DPLNNTPTVYGASPGGLRELOLSLFEILKGVLIORNPOLCYODITLMKDIFFHKNNOA 180
|||
Db 121 DPLNNTPTVYGASPGGLRELOLSLFEILKGVLIORNPOLCYODITLMKDIFFHKNNOA 180
|||
Qy 181 LTLIDTNRSRACHPCSPMKGSRMCWGESSEDCOSLTRTYCAGGCARCKGPLPTDCCHQC 240
|||
Db 181 LTLIDTNRSRACHPCSPMKGSRMCWGESSEDCOSLTRTYCAGGCARCKGPLPTDCCHQC 240
|||
```

```
Qy 241 AACCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
|||
Db 241 AACCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
|||
Qy 301 YNLTSDVSGCTLVCPHLHNOEYTAEDGTORCEKCSKPCARVCYGLMEHLREVRAVTSAN 360
|||
Db 301 YNLTSDVSGCTLVCPHLHNOEYTAEDGTORCEKCSKPCARVCYGLMEHLREVRAVTSAN 360
|||
Qy 356 LRMOPG--PAHVPYSLFRPSMDLVSAFYSLPLAPLSPTSVPI-----SPYVGRGPD 405
|||
Db 361 IOEFACCKKIFGSLAFPLPESFDGPASNT---APLOPEOLQVFEITLLETITGLYISAMPD 417
|||
Qy 406 --PDAAVAVNLSRYEG 419
|||
Db 418 SLFDLSVFQNLQYIRG 433
|||
```

## RESULT 10

```
US-09-811-115-3
; Sequence 3, Application US/09811115
; Patent No. US20020035736A1
; GENERAL INFORMATION:
; APPLICANT: Erickson, Sharon
; APPLICANT: Schwalli, Ralph
; APPLICANT: King, Kathleen
; TITLE OF INVENTION: HER-2 TRANSGENIC NON-HUMAN TUMOR MODEL
; FILE REFERENCE: GENE.034A
; CURRENT APPLICATION NUMBER: US/09/811,115
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/189,844
; PRIOR FILING DATE: 2000-03-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-811-115-3
```

```
Query Match      82.1%; Score 1878; DB 10; Length 1255;
Best Local Similarity 83.0%; Pred. No. 3.5e-130;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
```

```
Qy 1 MELAALCRMGILLALPPAASSTOVCTGDMKLRLPASPTHLDMLRHLYOGCOVVOGNTL 60
|||
Db 1 MELAALCRMGILLALPPAASSTOVCTGDMKLRLPASPTHLDMLRHLYOGCOVVOGNTL 60
|||
Qy 61 ELTYLPTNASLSFLDIOEVQGYVLAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
|||
Db 61 ELTYLPTNASLSFLDIOEVQGYVLAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
|||
Qy 121 DPLNNTPTVYGASPGGLRELOLSLFEILKGVLIORNPOLCYODITLMKDIFFHKNNOA 180
|||
Db 121 DPLNNTPTVYGASPGGLRELOLSLFEILKGVLIORNPOLCYODITLMKDIFFHKNNOA 180
|||
Qy 181 LTLIDTNRSRACHPCSPMKGSRMCWGESSEDCOSLTRTYCAGGCARCKGPLPTDCCHQC 240
|||
Db 181 LTLIDTNRSRACHPCSPMKGSRMCWGESSEDCOSLTRTYCAGGCARCKGPLPTDCCHQC 240
|||
Qy 241 AACCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
|||
Db 241 AACCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDFESMPNPEGRTYFGASCYTACP 300
|||
Qy 301 YNLTSDVSGCTLVCPHLHNOEYTAEDGTORCEKCSKPCARVCYGLMEHLREVRAVTSAN 360
|||
Db 301 YNLTSDVSGCTLVCPHLHNOEYTAEDGTORCEKCSKPCARVCYGLMEHLREVRAVTSAN 360
|||
Qy 356 LRMOPG--PAHVPYSLFRPSMDLVSAFYSLPLAPLSPTSVPI-----SPYVGRGPD 405
|||
Db 361 IOEFACCKKIFGSLAFPLPESFDGPASNT---APLOPEOLQVFEITLLETITGLYISAMPD 417
|||
Qy 406 --PDAAVAVNLSRYEG 419
|||
```



```

; TYPE: PRT
;
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: HER500*-rGM-CSF construct
US-09-821-883-4

```

Query Match	70.4%;	Score 1610.5;	DB 10;	Length 697;
Best Local Similarity	75.4%;	Pred. No. 8.1e-11;		
Matches 319; Conservative	8;	Mismatches 53;	Indels 43;	Gaps 7

QY	16	LPRFAASTVOSTGTDKMLRLPASETHILMDI,RHLYGOCQVVOGSLLETLYLPTNASTSLFQ	75
Dd	35	LARGAASTVOSTGTDKMLRLPASETHILMDI,RHLYGOCQVVOGSLLETLYLPTNASTSLFQ	94
QY	76	DIDVGVGYVLIAHQVQVPLQRLRYRGVQLDFEDNYALAVLDNGDPLNNTTPVGTGASPG	135
Dd	95	DIDVGVGYVLIAHQVQVPLQRLRYRGVQLDFEDNYALAVLDNGDPLNNTTPVGTGASPG	154
QY	136	GLRFLQRLSTLEILKGGVLLIQRNPOILCYOFTLLMKDIFHNNOALATLLIDTNSRACHPC	195
Dd	155	GLRFLQRLSTLEILKGGVLLIQRNPOILCYOFTLLMKDIFHNNOALATLLIDTNSRACHPC	214
QY	196	SPMKSGSRCKGESESDCOSLTRVCAGGACARCKGRLPTDCCHEQCAAGCTGPRKHSOLAC	255
Dd	215	SPMKSGSRCKGESESDCOSLTRVCAGGACARCKGRLPTDCCHEQCAAGCTGPRKHSOLAC	274
QY	256	LHFHNSICELCPALATYTTIDTFEESPRNDEGRYTFGASCVTACPYNYLSTDVGSCTLVC	315
Dd	275	LHFHNSICELCPALATYTTIDTFEESPRNDEGRYTFGASCVTACPYNYLSTDVGSASTI-	333
QY	316	PLHQEYTAEDGTCQCKGCKSPARG-----THSLPRPAVR--VPLMQSG-----	361
Dd	334	---NFEKLGAGGAVNNHNRSSSTRSGGDLTLGLSEFEERAPSPPLASFGAGSDVAFDGD	390
QY	362	-----PAHPVLSFLRPSPWLDVSAFYSLLPLPL--SPNSVSPISPVGVSGPD---	405
Dd	391	LGWGAAGKGLQSLTPHDSPLQRYRSED-----PVPRLRSESDGYAARLTCSPQREYVN	442
QY	406	-PD 407	
Dd	443	QPD 445	

```

RESULT 14
US-09-854-356-8
Sequence 8, Application US/09854356
Patent No. US20020177567A1
GENERAL INFORMATION:
APPLICANT: Cheever, Martin A.
APPLICANT: Cheyssen, Dirk
APPLICANT: Corixa Corporation
APPLICANT: SmithKline Beecham Biologicals S. A.
TITLE OF INVENTION: HER-2/neu Fusion Proteins
FILE REFERENCE: 014058-009810PC
CURRENT APPLICATION NUMBER: US/09/854,356
CURRENT FILING DATE: 2001-05-09
PRIOR APPLICATION NUMBER: US 09/493,480
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: US 60/117,976
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 8
LENGTH: 654
TYPE: PRT
ORGANISM: Rattus sp.
FEATURE:
OTHER INFORMATION: extracellular domain (ECD) of rat HER-2/neu
US-09-854-356-8

```

Query Match	70.3%;	Score 1608.5;	DB 9;	Length 654;
Best Local Similarity	85.0%;	Pred. No. 1.1e-110;		
Matches 294;	Conservative 15;	Mismatches 36;	Indels 1;	Gaps 1

QY	1	MEALACRMGILLLLLP	GAAS	TOSTG	TDKIL	ELPAS	PEH	NDML	HHV	IOG	COV	GNL	60
Db	1	MELAAKRMGILLLLP	PGIAG	TOSTG	TDKIL	ELPAS	PEH	NDML	HHV	IOG	COV	GNL	60
QY	61	ELATPLPNASISL	FDIOIE	VOGYV	LIAH	NOVR	QVPL	ORLRL	IVRG	TOL	FEDNY	ALAVLDNG	120
Db	61	ELATVVPANASIS	LFDIOIE	VOGYV	LIAH	NOYK	RPR	LDRLRL	IVRG	TOL	FEDKY	ALAVLDNR	120
QY	121	DPLNTPTPV	TGAS	PGGLREL	QLRSL	TEIL	KGV	LIO	RNP	OLCY	DTIL	MKDFIFHKNNOL	179
Db	121	DQONVAASPT	PGRTPEG	IREL	QLRSL	TEIL	KGV	LIG	NPQL	LCY	QDM	VLWKDFYFRKNNOL	180
QY	180	ALTLIDITNRS	ACHPC	SPMCK	KGSK	WESS	EDCOS	LTRY	CA	GCAN	CKGRL	PLPDCHEQ	239
Db	181	APVVIDITNRS	ACHPC	SPACK	DKNH	CMWESS	EDCO	ILIT	GT	SC	CAK	CKGRLPLPDCHEQ	240
QY	240	CAACCTGPKHSD	CLACH	FNHNS	GICEL	HCAL	YVYNN	DTFES	MPN	PNRG	RTFG	SCYTAC	299
Db	241	CAACCTGPKHSD	CLACH	FNHNS	GICEL	HCAL	YVYNN	DTFES	MPN	PNRG	RTFG	SCYTAC	300
QY	300	PYNLTSDVGSCT	LVCP	LDLH	NOVE	VAED	GTOR	CEK	CSK	PCAR	GTSH	345	
Db	301	PYNLTSTEVGSCT	LVCP	PPND	NOVE	VAED	GTOR	CEK	CSK	PCAR	AVCTGL	346	

RESULT 15  
 US-09-854-356-2  
 : Sequence 2, Application US/09854356  
 : Patent No. US20020177567A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Cheever, Martin A.  
 : APPLICANT: Gheysen, Dirk  
 : APPLICANT: Corixa Corporation  
 : APPLICANT: Smithkline Beecham Biologicals S. A.  
 : TITLE OF INVENTION: HER-2/neu Fusion Proteins  
 : FILE REFERENCE: 014058-009810PC  
 : CURRENT APPLICATION NUMBER: US/09/854,356  
 : CURRENT FILING DATE: 2001-05-09  
 : PRIOR APPLICATION NUMBER: US 09/493,480  
 : PRIOR FILING DATE: 2000-01-28  
 : PRIOR APPLICATION NUMBER: US 60/117,976  
 : PRIOR FILING DATE: 1999-01-29  
 : NUMBER OF SEQ ID NOS: 26  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 2  
 : LENGTH: 1256  
 : TYPE: PRT  
 : ORGANISM: Rattus sp.  
 : FEATURE:  
 : OTHER INFORMATION: rat HER-2/neu protein  
 : NAME/KEY: DOMAIN  
 : LOCATION: (1)..(654)  
 : OTHER INFORMATION: extracellular domain (ECD)  
 : NAME/KEY: DOMAIN  
 : LOCATION: (677)..(1256)  
 : OTHER INFORMATION: intracellular domain (ICD)  
 : NAME/KEY: DOMAIN  
 : LOCATION: (721)..(998)  
 : OTHER INFORMATION: kinase domain (KD)  
 : NAME/KEY: DOMAIN  
 : LOCATION: (991)..(1256)  
 : OTHER INFORMATION: phosphorylation domain (PD)  
 : NAME/KEY: DOMAIN  
 : LOCATION: (991)..(1049)  
 : OTHER INFORMATION: fragment of the phosphorylation domain, preferred  
 : :S-09-854-356-2

Query Match	70.3%;	Score 1608.5;	DB 9;	Length 1256;
Best Local Similarity	85.0%;	Pred. NO. 2.3e-110;		
Matches 294;	Conservative 15;	Mismatches 36;	Indels 1;	Gaps 1;

```
OY 1 MELALCRMGILLALLPPGAASTOVCTGTDMLRLPASPETHLDMLRHLYOGCQVVOGNL 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MELAMCRMGILLALLPPGIAGTQVCTGTDMLRLPASPETHLDMLRHLYOGCQVVOGNL 60
OY 61 ELTYLPTNASLSFLQDIOEVQGYVLIANOVROYPLQRLRIVRGTQLFEDNVALAVLDNG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ELTYVPANASLSFLQDIOEVQGYMLIANOVKRVPLQRLRIVRGTQLFEDKVALAVLDNR 120
OY 121 DPLNNTTPTVT-GASPGGLRELOLRSLTEILKGGVLIQBNPOLCYODTILMKDIFHKNNOL 179
    || : || : || : ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 DPQDNVAASTPGKTPESGLRELOLRSLTEILKGGVLIQBNPOLCYQDNVLMKDVFRKNOL 180
OY 180 ALTLIDTNSRACHPCSPMCKGSRGSESDQSLRTVCAGGACARCKGPLEPTDCHEQ 239
    | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 APYDIDTNSRACPPCAPACKDNHCWESPEDCQILGTICTSGCARCKGRLEPTDCHEQ 240
OY 240 CAAGCTGPKHSDCLACLHFHNSGICELHCPALVTYNTDTFESMPNPEGRTTFGASCVTAC 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 CAAGCTGPKHSDCLACLHFHNSGICELHCPALVTYNTDTFESMHNPEGRTTFGASCVTTC 300
OY 300 PNYLSTDVSGCTLVGCPLNQEVTAEDGTQRCCKSPCARVGYGL 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 PNYLSTEVSGCTLVGCPNNQEVTAEDGTQRCCKSPCARVGYGL 346
```

Search completed: March 4, 2003, 12:43:46  
Job time : 20.6687 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:36:41 ; Search time 10.1526 Seconds  
(without alignments)  
1036.856 Million cell updates/sec

Title: US-09-234-208B-1

Perfect score: 79

Sequence: 1 GTHSLPRPAAPVPLRMQP.....VGRGPDPAHVAVNLRYEG 79

Scoring table: OLIGO /  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A.Geneseq\_101002.\*

```
1: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	79	22	AAE09196 Human p68HER-2 ECD
2	79	100.0	83	22	AAE09185 Human p68HER-2 par
3	79	100.0	84	23	AAE20349 HER2 protein conta
4	73	92.4	79	22	AAE09195 Human p68HER-2 ECD
5	73	92.4	79	22	AAE09199 Human p68HER-2 ECD
6	72	91.1	79	22	AAE09188 Human p68HER-2 ECD
7	66	83.5	79	21	AAV97241 HER-2 C-terminal e
8	66	83.5	79	22	AAE09184 Human p68HER-2 ECD
9	66	83.5	79	22	AAE09186 Human p68HER-2 ECD
10	66	83.5	79	22	AAE09187 Human p68HER-2 ECD

11	66	83.5	79	23	AAE20350 Human HER2 Intron
12	57	72.2	79	22	AAE09194 Human p68HER-2 ECD
13	56	70.9	79	22	AAE09189 Human p68HER-2 ECD
14	55	69.6	79	22	AAE09197 Human p68HER-2 ECD
15	54	68.4	79	22	AAE09190 Human p68HER-2 ECD
16	51	64.6	79	22	AAE09191 Human p68HER-2 ECD
17	47	59.5	79	22	AAE09193 Human p68HER-2 ECD
18	41	51.9	79	22	AAE09198 Human p68HER-2 ECD
19	36	45.6	79	22	AAE09192 Human p68HER-2 ECD
20	17	21.5	79	21	AAV97239 HER-2 C-terminal p
21	17	21.5	79	22	AAE09180 Human p68HER-2 ECD
22	17	21.5	79	22	AAE09182 Human p68HER-2 ECD
23	17	21.5	79	23	AAE20347 Human HER2 Intron
24	15	19.0	419	22	AAE09212 Human p68HER-2 gen
25	15	19.0	419	22	AAE09213 Human p68HER-2 gen
26	15	19.0	419	22	AAE09216 Human p68HER-2 gen
27	14	17.7	419	22	AAE09181 Human p68HER-2 gen
28	14	17.7	419	22	AAE09200 Human p68HER-2 gen
29	14	17.7	419	22	AAE09203 Human p68HER-2 gen
30	14	17.7	419	22	AAE09204 Human p68HER-2 gen
31	14	17.7	419	22	AAE09205 Human p68HER-2 gen
32	14	17.7	419	22	AAE09206 Human p68HER-2 gen
33	14	17.7	419	22	AAE09207 Human p68HER-2 gen
34	14	17.7	419	22	AAE09208 Human p68HER-2 gen
35	14	17.7	419	22	AAE09209 Human p68HER-2 gen
36	14	17.7	419	22	AAE09210 Human p68HER-2 gen
37	14	17.7	419	22	AAE09211 Human p68HER-2 gen
38	14	17.7	419	23	AAE20348 Human truncated HE
39	14	17.7	420	21	AAV97240 Human p68HER-2, p
40	12	15.2	419	22	AAE09183 Human p68HER-2 gen
41	12	15.2	419	22	AAE09202 Human p68HER-2 gen
42	12	15.2	419	22	AAE09214 Human p68HER-2 gen
43	12	15.2	419	22	AAE09215 Human p68HER-2 gen
44	7	8.9	61	22	AAE09214 Human brain expres
45	7	8.9	61	22	AAE09214 Human bone marrow

## ALIGNMENTS

```
RESULT 1
AAE09196
ID AAE09196 standard; peptide; 79 AA.
AC AAE09196;
XX
DT 15-NOV-2001 (first entry)
XX
DE Human p68HER-2 ECDIIIA variant 11 encoded by HER-2 Intron 8.
DE
DE HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KW Solid tumour; Cancer; polymorphism; cytostatic; gene therapy;
KW p68HER-2; ECDIIIA; variant.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX MISC-difference 6 /note="p68HER-2 ECDIIIA (AAE09184) Pro substituted
XX FT with Ieu"
XX FT
XX FT MISC-difference 73 /note="p68HER-2 ECDIIIA (AAE09184) Asp substituted
XX FT with Asn which is encoded by CAC"
XX
XX WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
XX 16-FEB-2000; 2000US-0506079.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
```

```

XX Clinton G, Henner WD, Evans A;
PI
XX WPI: 2001-529934/58.
DR N-PSDB; AAD15865.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
PI the treatment of hard tumors -
XX
XX Example 11: Page -: 61pp; English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Herceptin, at an affinity of at least 10-8. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of Intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC the ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is p68HER-2 ECDIIIA variant
CC Note: The present sequence is not shown in the specification but is
CC derived from HER-2 Intron 8 encoded ECDIIIA sequence given in figure 8
CC (AAE09184).
XX
XX Sequence 79 AA:
SQ
XX
XX Query Match 100.0%; Score 79; DB 22; Length 79;
XX Best Local Similarity 100.0%; Pred. No. 2e-71;
XX Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GTHSLPRPAAVPVRLMQPGPAHVPFLRPSWDLVSFAFYSIPLAPISPTSPISPSV 60
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1 GTHSLPRPAAVPVRLMQPGPAHVPFLRPSWDLVSFAFYSIPLAPISPTSPISPSV 60
QY 61 GRGPDPDAHVAVNLRSYEG 79
DB 61 GRGPDPDAHVAVNLRSYEG 79
XX
XX RESULT 2
XX AAE09185 standard; Protein: 83 AA.
XX
XX AAE09185;
XX
XX 15-NOV-2001 (first entry)
XX
XX Human p68HER-2 partial protein containing ECDIIIA variant sequence.
XX
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
XX solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA; variant.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 3..81
XX /label= ECDIIIA
XX /note= "Extracellular domain IIIA of p68HER-2"
XX
XX Misc-difference 1 /note= "Encoded by CC"
XX
XX Misc-difference 8 /note= "Most commonly occurring ECDIIIA (AAE09184) Pro
XX substituted with Leu"
XX
XX Modified-site 75 /note= "Asn is N-glycosylated. Most commonly occurring
XX ECDIIIA (AAE09184) Asp substituted with Asn"
XX
XX

```

```

PN WO200161356-A1.
XX
XX 23-AUG-2001.
PD
XX
XX 16-FEB-2001; 2001WO-US05327.
PF
XX
XX 16-FEB-2000; 2000US-0506079.
PR
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
PA
XX
XX Clinton G, Henner WD, Evans A;
PI
XX WPI: 2001-529934/58.
DR N-PSDB; AAD15854.
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
PI the treatment of hard tumors -
XX
XX Example 1: Fig 1: 61pp; English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Herceptin, at an affinity of at least 10-8. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of Intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is human p68HER-2 partial protein
CC containing ECDIIIA variant sequence.
XX
XX Sequence 83 AA:
SQ
XX
XX Query Match 100.0%; Score 79; DB 22; Length 83;
XX Best Local Similarity 100.0%; Pred. No. 2.1e-71;
XX Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GTHSLPRPAAVPVRLMQPGPAHVPFLRPSWDLVSFAFYSIPLAPISPTSPISPSV 60
QY |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 3 GTHSLPRPAAVPVRLMQPGPAHVPFLRPSWDLVSFAFYSIPLAPISPTSPISPSV 62
QY 61 GRGPDPDAHVAVNLRSYEG 79
DB 63 GRGPDPDAHVAVNLRSYEG 81
XX
XX RESULT 3
XX AAE20349 standard; Protein: 84 AA.
XX
XX AAE20349;
XX
XX 18-JUN-2002 (first entry)
XX
XX HER2 protein containing extracellular domain (ECDIIIA).
XX
XX Endothelial growth factor receptor; EGFR; tumour; cytostatic; herstatin;
XX HER-2 receptor tyrosine kinase; squamous cell carcinoma; lung; colon;
XX glial cell tumour; cell growth.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
XX Misc-difference 1..2 /note= "Encoded by CCCGA"
XX
XX Modified-site 75..77 /note= "Asn is N-glycosylated"
XX
XX Misc-difference 81..82 /note= "Encoded by GGCTGAGACGGCCCTTCCGCCACCCACCCGCCAC
XX
XX

```

FT	TCCCTCAGTCG"
FT	Misc-difference 83..84
XX	/note= "Encoded by TIGCT"
XX	
PN	MO200214A70-A2.
PD	21-FEB-2002.
XX	
XX	14-AUG-2001; 2001WO-US25502.
PF	
XX	14-AUG-2000; 2000US-0638834.
PR	
XX	(UYOR-) UNIV OREGON HEALTH SCI.
PA	
XX	Clinton GM;
PI	
XX	WPI: 2002-269185/31.
DR	
XX	N-PSDB; AAD32539.
PT	
PT	Treating solid tumor characterized by expression of endothelial growth factor receptor, involves administering recombinant herstatin that binds to extracellular domain of the endothelial growth factor receptor
XX	
PS	Example 11; Fig 1; 82pp: English.
XX	
CC	The present invention relates to a method for treating a solid tumour characterised by endothelial growth factor receptor (EGFR) expression.
CC	The method involves administering an agent that binds to an extracellular domain (ECD) of EGFR. The invention also relates to a naturally occurring
CC	inhibitor of HER-2 receptor tyrosine kinase called herstatin. The co-expression of herstatin with p185HER2 causes a striking reduction in cell growth that corresponds with suppression of p185 autophosphorylation. The
CC	method or a pharmaceutical composition is useful for treating a solid tumour (selected from squamous cell carcinoma, lung carcinoma, colon carcinoma and gtrial cell tumour) characterised by EGFR expression. The
CC	present sequence is HER2 protein containing extracellular domain (ECDIIIIa) sequence.
CC	
XX	
SQ	Sequence 84 AA:
	Query Match 100.0%; Score 79; DB 23; Length 84;
	Best Local Similarity 100.0%; Pred. No. 2,1e-71;
	Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GTHSLPRPAVPPVPLRMQGPDAHPVLSFLRPSMVLVSATFSLPLAPISPTSPVSPVSV 60
DB	3 GTHSLPRPAVPPVPLRMQGPDAHPVLSFLRPSMVLVSATFSLPLAPISPTSPVSPVSV 62
QY	61 GRGPPDAHVAVNLSRYEG 79
DB	63 GRGPPDAHVAVNLSRYEG 81
RESULT 4	
AAE09195	
ID	AAE09195 standard; peptide: 79 AA.
XX	
AC	AAE09195;
XX	
DT	15-NOV-2001 (first entry)
XX	
DE	Human p68HER-2 ECDIIIIa variant 10 encoded by HER-2 Intron 8.
XX	
KM	HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KW	solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX	p68HER-2; ECDIIIIa; variant.
XX	
OS	Homo sapiens.
XX	
XX	
Key	Location/Qualifiers
FT	Misc-difference 73
FT	/note= "p68HER-2 ECDIIIIa (AAE09184) ASP substituted

[illegible]

```
FT /note= "p68HER-2 ECDIIIA (AAE09184) Asp substituted
FT with Asn"
XX
XX WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
XX 16-FEB-2000; 2000US-0506079.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Clinton G, Henner WD, Evans A;
XX
XX WPI: 2001-529934/58.
XX
XX N-PSDB: AAD15868.
XX
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
XX the treatment of hard tumors -
XX
XX Example 12: Page -: 61pp; English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
XX particularly a polypeptide that binds to the extracellular domain (ECD)
XX of HER-2 at a site that is different from the binding site of humanised
XX antibody, Herceptin, at an affinity of at least 1078. The present
XX invention is based upon the initial discovery of an alternative HER-2
XX mRNA transcript with 274 bp insert of intron 8. The translation product
XX of the alternative transcript is a truncated HER-2 protein designated
XX p68HER-2 which lacks the transmembrane and intracellular domains of
XX p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
XX The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
XX the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
XX nucleic acids encoding these are useful to treat, diagnose and identify
XX solid tumours. The present sequence is p68HER-2 ECDIIIA variant
XX encoded by polymorphic form of human HER-2 intron 8.
XX Note: The present sequence is not shown in the specification but is
XX derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
XX (AAE09184).
XX
XX Sequence 79 AA:
SQ
Query Match 92.4%; Score 73; DB 22; Length 79;
Best Local Similarity 100.0%; Pred. No. 2e-65;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 PRPAVPVPLRMQPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPVSGRGGPP 66
DB 7 PRPAVPVPLRMQPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPVSGRGGPP 66
QY 67 DAHVAVNLSRYEG 79
DB 67 DAHVAVNLSRYEG 79
RESULT 6
AAE09188
ID AAE09188 standard; peptide; 79 AA.
XX
XX AAE09188;
XX
XX 15-NOV-2001 (first entry)
XX
XX Human p68HER-2 ECDIIIA variant 3 encoded by HER-2 intron 8.
XX
XX HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
XX solid tumour; cancer; polymorphism; cytostatic; gene therapy;
XX p68HER-2; ECDIIIA; variant.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
```

```
FT Misc-difference 6
FT /note= "p68HER-2 ECDIIIA (AAE09184) Pro substituted
FT with Leu"
XX
XX WO200161356-A1.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
XX
XX 16-FEB-2000; 2000US-0506079.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Clinton G, Henner WD, Evans A;
XX
XX WPI: 2001-529934/58.
XX
XX N-PSDB: AAD15857.
XX
XX
XX New polypeptide, which binds to the extracellular domain of HER-2 for
XX the treatment of hard tumors -
XX
XX Example 11: Page -: 61pp; English.
XX
XX The invention relates to novel HER-2 (herstatin-2) antagonist
XX particularly a polypeptide that binds to the extracellular domain (ECD)
XX of HER-2 at a site that is different from the binding site of humanised
XX antibody, Herceptin, at an affinity of at least 1078. The present
XX invention is based upon the initial discovery of an alternative HER-2
XX mRNA transcript with 274 bp insert of intron 8. The translation product
XX of the alternative transcript is a truncated HER-2 protein designated
XX p68HER-2 which lacks the transmembrane and intracellular domains of
XX p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.
XX The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise
XX the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
XX nucleic acids encoding these are useful to treat, diagnose and identify
XX solid tumours. The present sequence is p68HER-2 ECDIIIA variant
XX encoded by polymorphic form of human HER-2 intron 8.
XX Note: The present sequence is not shown in the specification but is
XX derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8
XX (AAE09184).
XX
XX Sequence 79 AA:
SQ
Query Match 91.1%; Score 72; DB 22; Length 79;
Best Local Similarity 100.0%; Pred. No. 2e-64;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTHSLPRPAVPVPLRMQPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPVSV 60
DB 1 GTHSLPRPAVPVPLRMQPGPAHPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPVSV 60
QY 61 GRGPPDAHVAV 72
DB 61 GRGPPDAHVAV 72
RESULT 7
AAV97241
ID AAV97241 standard; Protein; 79 AA.
XX
XX AAV97241;
XX
XX 04-DEC-2000 (first entry)
XX
XX HER-2 C-terminal extracellular domain IIIA.
XX
XX HER-2; erbB-2; oncogene; receptor-like tyrosine kinase; insertion;
XX extracellular domain IIIA; antagonist; intron 8; C-terminal extension;
XX truncated HER-2; p68; dimerization inhibitor; cytostatic.
XX
XX Homo sapiens.
XX
```



```
XX 16-FEB-2001; 2001WO-US05327.
PF 16-FEB-2001; 2001WO-US05327.
XX 16-FEB-2000; 2000US-0506079.
PR 16-FEB-2000; 2000US-0506079.
XX (UYOR-) UNIV OREGON HEALTH SCI.
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX Clinton G, Henner WD, Evans A;
PI Clinton G, Henner WD, Evans A;
XX WPI: 2001-529934/58.
DR N-PSDB: AAD15855.
XX
PT New polypeptide, which binds to the extracellular domain of HER-2 for
PT the treatment of hard tumors -
XX
XX Example 11: Page -: 61pp: English.
PS
XX The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Herceptin, at an affinity of at least 108. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of Intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIa.
CC The ECDIIa-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is p68HER-2 ECDIIa variant
CC encoded by polymorphic form of human HER-2 intron 8.
CC Note: The present sequence is not shown in the specification but is
CC derived from HER-2 intron 8 encoded ECDIIa sequence given in figure 8
CC (AAE09184).
XX
SQ Sequence 79 AA:
Query Match 83.5%; Score 66; DB 22; Length 79;
Best Local Similarity 100.0%; Pred. No. 2e-58;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 PRPAVPVPLRMQPGPAHVLSTFLRPSMDVSAFYSLPLAPLSPVSPISPVSGRGDP 66
Db 7 PRPAVPVPLRMQPGPAHVLSTFLRPSMDVSAFYSLPLAPLSPVSPISPVSGRGDP 66
QY 67 DAHAVAV 72
Db 67 DAHAVAV 72
Db 67 DAHAVAV 72
RESULT 10
AAE09187
ID AAE09187 standard; peptide; 79 AA.
XX
AC AAE09187;
XX
DT 15-NOV-2001 (first entry)
XX
DE Human p68HER-2 ECDIIa variant 2 encoded by HER-2 intron 8.
XX
KW HER-2; herstatin; antagonist; extracellular domain; ECD; Herceptin;
KW solid tumour; cancer; polymorphism; cytostatic; gene therapy;
KW p68HER-2; ECDIIa; variant.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 5
FT /note="p68HER-2 ECDIIa (AAE09184) Ieu substituted
FT with Pro"
XX
XX WO200161356-A1.
XX
```

```
PD 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05327.
PF 16-FEB-2001; 2001WO-US05327.
XX 16-FEB-2000; 2000US-0506079.
PR 16-FEB-2000; 2000US-0506079.
XX (UYOR-) UNIV OREGON HEALTH SCI.
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX Clinton G, Henner WD, Evans A;
PI Clinton G, Henner WD, Evans A;
XX WPI: 2001-529934/58.
DR N-PSDB: AAD15856.
XX
PT New polypeptide, which binds to the extracellular domain of HER-2 for
PT the treatment of hard tumors -
XX
XX Example 11: Page -: 61pp: English.
PS
XX The invention relates to novel HER-2 (herstatin-2) antagonist
CC particularly a polypeptide that binds to the extracellular domain (ECD)
CC of HER-2 at a site that is different from the binding site of humanised
CC antibody, Herceptin, at an affinity of at least 108. The present
CC invention is based upon the initial discovery of an alternative HER-2
CC mRNA transcript with 274 bp insert of Intron 8. The translation product
CC of the alternative transcript is a truncated HER-2 protein designated
CC p68HER-2 which lacks the transmembrane and intracellular domains of
CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIa.
CC The ECDIIa-containing polypeptides bind tightly to, and thus antagonise
CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the
CC nucleic acids encoding these are useful to treat, diagnose and identify
CC solid tumours. The present sequence is p68HER-2 ECDIIa variant
CC encoded by polymorphic form of human HER-2 intron 8.
CC Note: The present sequence is not shown in the specification but is
CC derived from HER-2 intron 8 encoded ECDIIa sequence given in figure 8
CC (AAE09184).
XX
SQ Sequence 79 AA:
Query Match 83.5%; Score 66; DB 22; Length 79;
Best Local Similarity 100.0%; Pred. No. 2e-58;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 PRPAVPVPLRMQPGPAHVLSTFLRPSMDVSAFYSLPLAPLSPVSPISPVSGRGDP 66
Db 7 PRPAVPVPLRMQPGPAHVLSTFLRPSMDVSAFYSLPLAPLSPVSPISPVSGRGDP 66
QY 67 DAHAVAV 72
Db 67 DAHAVAV 72
Db 67 DAHAVAV 72
RESULT 11
AAE20350
ID AAE20350 standard; Protein; 79 AA.
XX
AC AAE20350;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human HER2 intron 8 encoded protein.
XX
KW Human; tumour; endothelial growth factor receptor; EGFR; cytostatic;
KW herstatin; HER-2 receptor tyrosine kinase; squamous cell carcinoma; Lung;
KW colon; glial cell tumour; cell growth.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 5
FT /note="p68HER-2 ECDIIa (AAE09184) Ieu substituted
FT with Pro"
XX
XX 14-AUG-2001; 2001WO-US25502.
XX
```







PS Example 11; Page -: 61pp; English.

XX The invention relates to novel HER-2 (herstatin-2) antagonist  
 CC particularly a polypeptide that binds to the extracellular domain (ECD)  
 CC of HER-2 at a site that is different from the binding site of humanised  
 CC antibody, Hereceptin, at an affinity of at least 10<sup>-8</sup>. The present  
 CC invention is based upon the initial discovery of an alternative HER-2  
 CC mRNA transcript with 274 bp insert of intron 8. The translation product  
 CC of the alternative transcript is a truncated HER-2 protein designated  
 CC p68HER-2 which lacks the transmembrane and intracellular domains of  
 CC p185HER-2 but contains ECD I, II of the p185HER-2 and the novel ECDIIIA.  
 CC The ECDIIIA-containing polypeptides bind tightly to, and thus antagonise  
 CC the HER-2 receptor. The peptides, which bind to an HER-2 ECD, and the  
 CC nucleic acids encoding these are useful to treat, diagnose and identify  
 CC solid tumours. The present sequence is p68HER-2 ECDIIIA variant  
 CC encoded by polymorphic form of human HER-2 intron 8.  
 CC Note: The present sequence is not shown in the specification but is  
 CC derived from HER-2 intron 8 encoded ECDIIIA sequence given in figure 8  
 CC (AAE09184).  
 CC  
 XX

SQ Sequence 79 AA;

Query Match 68.4%; Score 54; DB 22; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-46;  
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 OPQPAHPVLSFLRPSMDLYSAFYSLPLAPLSPVSPISPVSGRGPPDAHVAV 72  
 DB 19 QPEPAHPVLSFLRPSMDLYSAFYSLPLAPLSPVSPISPVSGRGPPDAHVAV 72

Search completed: March 4, 2003, 12:44:53  
 Job time : 10.1526 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:43:27 ; Search time 8.88353 Seconds  
(without alignments)  
854,910 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 79

Sequence: 1 GTHSLPRPAAVPVLRMQP.....VGRGPDPAHVAVNLSEYEG 79

Scoring table: OLIGO  
Gapop 60.0 , Gapept 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	10.1	520	2 C70311	hypothetical prote
2	8	10.1	865	2 A85032	hypothetical prote
3	8	10.1	879	2 S49910	chloroplast outer
4	8	10.1	1503	2 T01098	chloroplast outer
5	7	8.9	121	2 AH1848	hypothetical prote
6	7	8.9	189	2 S04670	hypothetical prote
7	7	8.9	207	2 A69941	capsular polysach
8	7	8.9	269	1 D43706	serine O-acetyltra
9	7	8.9	292	2 S40979	hypothetical prote
10	7	8.9	328	2 G84826	hypothetical prote
11	7	8.9	382	2 B88561	protein F58A4_7b f
12	7	8.9	389	2 A97577	hypothetical prote
13	7	8.9	389	2 AH2797	conserved hypotet
14	7	8.9	397	2 C83470	hypothetical prote
15	7	8.9	436	2 B70473	protoporphyrinogen
16	7	8.9	464	2 AF2180	diaminopimelate de
17	7	8.9	495	2 T17478	hypothetical prote
18	7	8.9	555	2 A26639	regulatory protein
19	7	8.9	574	2 S01272	regulatory protein
20	7	8.9	595	2 T15862	hypothetical prote
21	7	8.9	840	2 T36175	probable large ATP
22	7	8.9	895	2 A45554	thophy complex po
23	7	8.9	906	1 IJM5CN	N-cadherin precurs
24	7	8.9	2240	2 T37057	probable multi-dom
25	6	7.6	59	2 T07432	photosystem II pro
26	6	7.6	65	2 C72558	hypothetical prote
27	6	7.6	74	2 F82764	hypothetical prote
28	6	7.6	94	2 H81190	hypothetical prote
29	6	7.6	98	2 S68136	NADH2 dehydrogenas

30	6	7.6	101	2 H72493	hypothetical prote
31	6	7.6	119	2 S75551	hypothetical prote
32	6	7.6	120	2 D83166	hypothetical prote
33	6	7.6	122	2 AB2199	hypothetical prote
34	6	7.6	126	2 S25104	MSF14 protein - ma
35	6	7.6	128	2 G71264	conserved hypotet
36	6	7.6	128	2 A87510	hypothetical prote
37	6	7.6	129	2 B83173	hypothetical prote
38	6	7.6	142	2 C58723	hypothetical prote
39	6	7.6	142	2 E72571	hypothetical prote
40	6	7.6	145	1 J00347	capsid protein VP3
41	6	7.6	150	2 A69289	conserved hypotet
42	6	7.6	159	2 H70072	conserved hypotet
43	6	7.6	163	2 AB2025	hypothetical prote
44	6	7.6	171	2 T43959	hypothetical prote
45	6	7.6	182	2 S76346	hypothetical prote

## ALIGNMENTS

```
RESULT 1
C70311
hypothetical protein aq_116 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jun-2000
C:Accession: C70311
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300, MUID:98196666; PMID:9537320
A:Accession: C70311
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-520 <AOF>
A:Cross-references: GB:AE000674; NID:g2982850; PIDN:AA06482.1; PID:g2982861; GB:AE00
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_116
C:Superfamily: Aquifex aeolicus hypothetical protein aq_116

Query Match          10.1%; Score 8; DB 2; Length 520;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 PLSPSTVP 54
Db 2 PLSPSTVP 9

RESULT 2
A85032
hypothetical protein AT4902510 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold SP
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: A85032
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-865 <STO>
A:Cross-references: GB:NC_001268; NID:g7269011; PIDN:CAB80744.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4902510
A:Map position: 4

Query Match          10.1%; Score 8; DB 2; Length 865;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 9 PAAVPVPL 16  
|||||||  
Db 584 PAAVPVPL 591

RESULT 3  
S49910  
chloroplast outer envelope protein OEP86 precursor - garden pea  
N:Alternate names: chloroplast import-associated protein IAP86, GTP-binding  
C:Species: Pisum sativum (garden pea)  
C>Date: 26-Jan-1995 #sequence\_revision 10-Feb-1995 #text\_change 02-Feb-2001  
C:Accession: S49910; A53386; A55171  
R:Hirsch, S.; Soll, J.  
submitted to the EMBL Data Library, March 1994  
A:Reference number: S49910  
A:Accession: S49910

A:Molecule type: mRNA  
A:Residues: 1-879 <HIR>  
A:Cross-references: EMBL:231581; NID:959957; PID:959958  
R:Hirsch, S.; Muckel, E.; Heemeyer, F.; von Heljné, G.; Soll, J.  
Science 266, 1989-1992, 1994  
A:Title: A receptor component of the chloroplast protein translocation machinery.  
A:Reference number: A53386; M0ID:9509324; PMID:7801125  
A:Accession: A53386  
A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA  
A:Residues: 1-137, 0', 139-875, 877-879 <HI2>  
R:Kesseler, F.; Blobel, G.; Patel, H.A.; Schnell, D.J.  
Science 266, 1035-1039, 1994  
A:Title: Identification of two GTP-binding proteins in the chloroplast protein import ma  
A:Reference number: A55171; M0ID:95063938; PMID:7973656  
A:Accession: A55171  
A:Molecule type: mRNA  
A:Residues: 1-377, 'EOO', 376-879 <KE5>  
A:Cross-references: GB:U36657; NID:9576508; PID:NAA53276.1; PID:9576509  
C:Keywords: chloroplast; GTP binding; membrane protein; nucleotide binding; P-loop  
F:245-252/Region: nucleotide-binding motif A (P-loop)

Query Match 10.1%; Score 8; DB 2; Length 879;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 PAAVPVPL 16  
|||||||  
Db 599 PAAVPVPL 606

RESULT 4  
T01098  
chloroplast outer envelope protein OEP86 homolog T10P11.19 - Arabidopsis thaliana  
N:Alternate names: chloroplast import-associated protein, GTP-binding; protein T14P8.24  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 02-Feb-2001  
C:Accession: T01098; T01299  
R:Kaplan, N.; Johnson, D.; Schütz, K.; Gnoj, L.; Hoffman, J.; Tili, S.; de la Bastide, M.  
hi, M.; Martienssen, R.; Chen, E.Y.; Wilson, R.; McCombie, W.R.  
submitted to the EMBL Data Library, November 1998  
A:Description: Sequence of A. thaliana BAC T10P11 from chromosome IV.  
A:Reference number: Z14248  
A:Accession: T01098  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1503 <KAP>  
A:Cross-references: EMBL:AC002330; NID:92262135; PID:g3892053  
A:Experimental source: cultivar Columbia  
R:Kallick, J.; Elliott, G.; Cloud, J.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of A. thaliana T14P8.  
A:Reference number: Z14290  
A:Accession: T01299  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A:Residues: 1-1503 <KAL>  
A:Cross-references: EMBL:AF069298; NID:93193282; PID:g3193301  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 22/1  
A>Note: T10P11.19; T14P8.24  
C:Keywords: chloroplast; GTP binding; membrane protein; nucleotide binding; P-loop  
F:862-869/Region: nucleotide-binding motif A (P-loop)

Query Match 10.1%; Score 8; DB 2; Length 1503;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 PAAVPVPL 16  
|||||||  
Db 1222 PAAVPVPL 1229

RESULT 5  
AH1848  
hypothetical protein all0337 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AH1848  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Matanabe, A.; Iri  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 6, 205-213, 2001  
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AH1807; M0ID:21595285; PMID:11759840  
A:Accession: AH1848  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-121 <KUR>  
A:Cross-references: GB:BA000019; PID:BA072295.1; PID:g17129682; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all0337

Query Match 8.9%; Score 7; DB 2; Length 121;  
Best Local Similarity 100.0%; Pred. No. 8.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 PVLSPFLR 31  
|||||||  
Db 56 PVLSPFLR 62

RESULT 6  
S04670  
hypothetical protein 5 - Rhodospseudomonas blautica  
C:Species: Rhodospseudomonas blautica  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 18-Jun-1993  
C:Accession: S04670  
R:Tybuliewicz, V.L.J.; Falk, G.; Walker, J.E.  
J. Mol. Biol. 179, 185-214, 1984  
A:Title: Rhodospseudomonas blautica atp operon. Nucleotide sequence and transcription.  
A:Reference number: S04666; M0ID:85058188; PMID:6209404  
A:Accession: S04670  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-189 <TYB>

Query Match 8.9%; Score 7; DB 2; Length 189;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 PTSPVPS 56  
|||||||  
Db 84 PTSPVPS 90

```
RESULT 7
A:Accession: A69941
capsular polysaccharide homolog yypq - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: A69941
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle
Rieger, M.; Rivolta, C.; Rooha, E.; Roche, B.; Rose, M.; Sadleir, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
Aeuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69941
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-207 <KUN>
A:Cross-references: GB:299115; GB:AL009126; NID:92634478; PIDN:CAH14085.1; PID:ell83614;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yypq

Query Match
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LPRPAV 12
| | | | | |
Db 125 LPRPAV 131

RESULT 8
serine O-acetyltransferase (EC 2.3.1.30) nifp - Azotobacter chroococcum
C:Species: Azotobacter chroococcum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 05-May-2000
C:Accession: D43706
R:Evans, D.J.; Jones, R.; Woodley, P.R.; Wilborn, J.R.; Robson, R.L.
J. Bacteriol. 173, 5457-5469, 1991
A:Title: Nucleotide sequence and genetic analysis of the Azotobacter chroococcum nifusw
A:Reference number: A43706; MUID:91358323; PMID:1885524
A:Accession: D43706
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <EVA>
A:Cross-references: EMBL:M60090; NID:g142386; PIDN:AAA22162.1; PID:g142390
C:Superfamily: Bacillus serine acetyltransferase; serine acetyltransferase homology
C:Keywords: acyltransferase; coenzyme A
F:10-169/Domain: serine acetyltransferase homology <SAT>

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LPRPAV 12
| | | | | |
Db 52 LPRPAV 58

RESULT 9
hypochemical protein F58A.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 23-Feb-1997
```

```
C:Accession: S40979
R:Berts, M.
submitted to the EMBL Data Library, February 1992
A:Reference number: S40973
A:Accession: S40979
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-292 <BER>
A:Cross-references: EMBL:D22179
C:Genetics:
A:Insertions: 60/2; 133/1; 273/1
C:Keywords: nucleus

Query Match
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 APLSPTS 52
| | | | | |
Db 48 APLSPTS 54

RESULT 10
hypochemical protein At2g40220 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: G84826
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617137
A:Accession: G84826
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-328 <STO>
A:Cross-references: GB:AE002093; NID:g6598941; PIDN:AAF18736.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g40220
A:Map position: 2

Query Match
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLPRPA 11
| | | | | |
Db 144 LLPRPA 150

RESULT 11
B88561
protein F58A.7b [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: B88561
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: B88561
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <STO>
A:Cross-references: GB:chr_III; PIDN:CAA80170.1; PID:g3877829; GSPDB:GN00021; CESP:F5
C:Genetics:
A:Gene: F58A.7b
A:Map position: 3
```

Query Match 8.9%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 APLSPTS 52  
|||||||  
DB 48 APLSPTS 54

## RESULT 12

hypothetical protein AGR\_C3311 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: A97577  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: A97577  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-389 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK87570.1; PID:q15156910; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_3311  
A:Map position: circular chromosome

Query Match 8.9%; Score 7; DB 2; Length 389;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 PAAPVP 15  
|||||||  
DB 160 PAAPVP 166

## RESULT 13

conserved hypothetical protein Atu1800 [imported] - Agrobacterium tumefaciens (strain C5  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AH2797  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AH2797  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-389 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AAL42798.1; PID:q17740243; GSPDB:GN00186  
C:Genetics:  
A:Gene: Atu1800  
A:Map position: circular chromosome

Query Match 8.9%; Score 7; DB 2; Length 389;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 PAAPVP 15  
|||||||  
DB 160 PAAPVP 166

## RESULT 14

C83470

hypothetical protein PA1412 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C83470

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Ladbis, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: C83470

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-397 <STO>  
A:Cross-references: GB:AE004570; GB:AE004091; NID:9947350; PIDN:AAQ04801.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA1412

Query Match 8.9%; Score 7; DB 2; Length 397;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LLRPAA 11  
|||||||  
DB 187 LLRPAA 193

## RESULT 15

B70473  
protoporphyrinogen oxidase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C:Accession: B70473  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: B70473  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-436 <AOE>  
A:Cross-references: GB:AE000768; NID:92984249; PIDN:AAQ07778.1; PID:92984251; GB:AE00  
A:Experimental source: strain VP5  
C:Genetics:  
A:Gene: hemc

Query Match 8.9%; Score 7; DB 2; Length 436;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 VLSFLRP 32  
|||||||  
DB 428 VLSFLRP 434

Search completed: March 4, 2003, 12:47:41  
Job time : 10.8835 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:37:16 ; Search time 5.39357 Seconds  
(without alignments)  
607.506 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 79

Sequence: 1 GTHSLPRAAVPVPLRMQP.....VGRCPDPAHVAVNLRYEG 79

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	8.9	114	GON2_TUPGB	095336 tupala glis
2	7	8.9	189	YAN5_RHOBL	P05448 rhodopsendo
3	7	8.9	207	YQPQ_BACSU	P54183 bacillus su
4	7	8.9	269	NIFP_AZCH	P23145 azotobacter
5	7	8.9	380	YMH7_CAEEL	P34474 caenorhabdi
6	7	8.9	499	CPM1_ONCMY	092088 oncorhynch
7	7	8.9	574	ZEST_DROME	P09556 drosophila
8	7	8.9	635	SL56_HUMAN	P09289 homo sapien
9	7	8.9	677	GCP3_MOUSE	P58854 mus musculi
10	7	8.9	906	CAD2_MOUSE	P15116 mus musculi
11	7	8.9	907	GCP3_HUMAN	0966w5 homo sapien
12	7	8.9	968	ST10_HUMAN	094804 homo sapien
13	7	7.6	56	PSBK_PINTA	P41598 pinus thund
14	6	7.6	111	YCA5_OENHO	09amn3 oenothera h
15	6	7.6	126	MF14_MALIZ	001900 zea mays (m
16	6	7.6	141	RL9_HABCO	002376 haemochus
17	6	7.6	145	VP3_BPCHP	P19194 bacterioph
18	6	7.6	159	YXHD_BACSU	P46328 bacillus su
19	6	7.6	195	GSP1_ECOLI	P45761 escherichia
20	6	7.6	198	LMBV_CHICK	001636 gallus gall
21	6	7.6	204	LAFA_MOUSE	P51827 mus musculi
22	6	7.6	205	ESMC_DROME	001070 drosophila
23	6	7.6	206	COX3_BACFI	004442 bacillus fi
24	6	7.6	209	ADPP_ECOLI	P36651 escherichia
25	6	7.6	214	VAPD_DESSY	006506 desulfuroco
26	6	7.6	215	CRB6_ODOSI	P49488 odontella s
27	6	7.6	220	ACPD_STRCO	09slu6 streptomyce
28	6	7.6	220	BIOD_AOUAE	066832 atreflex aeo
29	6	7.6	235	GCH1_SCHPO	013774 schizosacch
30	6	7.6	239	PRR4_BOVIN	P18917 bos taurus
31	6	7.6	241	LAT_RAT	070601 ratius norv
32	6	7.6	242	LAT_MOUSE	054957 mus musculi
33	6	7.6	248	PT16_LYCCE	004682 lycopersico

34	6	7.6	250	1	MTR3_YEAST	P48240 saccharomyc
35	6	7.6	251	1	Y4EP_RHISN	P55429 rhizobium s
36	6	7.6	260	1	YH20_PASMU	09c425 pasteurella
37	6	7.6	263	1	COO4_CAEEL	P91428 caenorhabdi
38	6	7.6	290	1	AR11_MESAU	P50292 mesocricetu
39	6	7.6	290	1	MPN_HUMAN	09bq3 homo sapien
40	6	7.6	295	1	SSO2_YEAST	P39926 saccharomyc
41	6	7.6	301	1	MPEV_SYNPY	002178 synchococc
42	6	7.6	301	1	VF34_SYNPY	P74220 synchocyst
43	6	7.6	303	1	RDGC_ECO57	08x6al escherichia
44	6	7.6	303	1	RDGC_ECOLI	P36767 escherichia
45	6	7.6	303	1	RDGC_SALTY	08zre6 salmoneella

## ALIGNMENTS

```

RESULT 1
ID GON2_TUPGB          STANDARD:      PRT: 114 AA.
AC 095336:
DT 15-DEC-1998 (rel. 37, Created)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE Progonadolibertin II precursor [Contains: Gonadolibertin II (LHRH II)
DE (luteinizing hormone releasing hormone II) (Gonadotropin releasing
DE hormone II) (GNRH II) (Lulibertin II); GNRH-associated peptide II].
GN GNRH2.
OS Tupala glis belangeri (Common tree shrew).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Scandentia; Tupaiidae; Tupata.
OX NCBI_TaxID=9396;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hypothalamus;
RX MEDLINE=97079639; PubMed=8921350;
RA Kasten T.L., White S.A., Norton T.T., Bond C.T., Adelman J.P.,
RA Fernald R.D.;
RT *Characterization of two new preproGNRH mRNAs in the tree shrew:
RT first direct evidence for mesencephalic GNRH gene expression in a
RT placental mammal.
RT Gen. Comp. Endocrinol. 104:7-19(1996).
CC -!- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS; IT STIMULATES
CC THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING
CC HORMONES.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: MIDBRAIN.
CC -!- SIMILARITY: BELONGS TO THE GNRH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U63327; AAB1638.1; .
DR InterPro: IPR002012; GNRH.
DR Pfam: PF00446; GNRH.1.
DR PROSITE: PS00473; GNRH.1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Signal.
FT SIGNAL 1 25
FT CHAIN 26 114 PROGNADOLIBERTIN II.
FT PEPTIDE 26 35 GNRH-ASSOCIATED PEPTIDE II.
FT MOD_RES 39 114 PYRROLIDONE CARBOXYLIC ACID
FT MOD_RES 26 26 (BY SIMILARITY).
FT MOD_RES 35 35 AMIDATION (G-36 PROVIDE AMIDE GROUP).
SQ SEQUENCE 114 AA: 12123 MW: 680690ELC6869BEC1 CXC64;
Query Match 8.9%; Score 7; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 4.2;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 PRPAAP 13  
 |||||  
 Db 105 PRPAAP 111

## RESULT 2

YAF5\_RHOBL STANDARD: PRT: 189 AA.  
 ID YAF5\_RHOBL  
 AC P05448;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-NOV-1988 (Rel. 09, Last annotation update)  
 DE ATP synthase subunits region ORF 5.  
 OS Rhodospirillum rubrum.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;  
 CC Rhodospirillum.  
 OX NCBI\_TaxID=1075;  
 RN [1]  
 RP MEDLINE=85058188; PubMed=6209404;  
 RX Tyndale V.L.J., Falk G., Walker J.E.;  
 RT "Rhodospirillum rubrum atp operon. Nucleotide sequence and  
 transcription.";  
 RL J. Mol. Biol. 179:185-214(1984).

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: 200018; CAA77311.1;  
 DR PIR: S04670; S04670.  
 KW Hypothetical protein.  
 SQ SEQUENCE 189 AA; 20759 MW; 68756B3A3C8C466 CRC64;

Query Match 8.9%; Score 7; DB 1; Length 189;  
 Best Local Similarity 100.0%; Pred. No. 6.6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 50 PRSPIS 56  
 |||||  
 Db 84 PRSPIS 90

RESULT 3  
 YPOP\_BACSU STANDARD: PRT: 207 AA.  
 ID YPOP\_BACSU  
 AC P54183;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ypop (Fragment).  
 GN ypop.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / Marburg;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azavedo V., Zumbstein E., Galleron N., Ehrlich S.D.,  
 RA Serron P.;  
 RT "Sequence analysis of the Bacillus subtilis chromosome region between  
 RT the sera and kag loci cloned in a yeast artificial chromosome.";  
 RL Microbiology 142:2005-2016(1996).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 OC Azotobacter.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: L77246; AAA96649.1;  
 DR EMBL: 299115; CAB14085.1;  
 DR Subtilisin; BGL1629; YPQP.  
 DR Interpro: IPR003869; Polysac\_Cap.  
 DR Pfam: PF02719; Polysac\_syn\_2; 1.  
 KW Hypothetical protein; Complete proteome.  
 FT NON\_TER  
 SQ SEQUENCE 207 AA; 23286 MW; F3EA64945DE249CD CRC64;

Query Match 8.9%; Score 7; DB 1; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 7.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 LPPPAV 12  
 |||||  
 Db 125 LPPPAV 131

## RESULT 4

NIFP\_AZCOH STANDARD: PRT: 269 AA.  
 ID NIFP\_AZCOH  
 AC P23145;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE Probable serine acetyltransferase (EC 2.3.1.30) (SAT).  
 GN NIFP.  
 OS Azotobacter chroococcum mcd 1.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Azotobacter.



```

OX NCBI_TaxID=355;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=91358323; PubMed=1885524;
RA Evans D.J., Jones R., Woodley P.R., Wilborn J.R., Robson R.L.;
RT "Nucleotide sequence and genetic analysis of the Azotobacter
RT chroococcum nifUSVWZ gene cluster, including a new gene (nifP) which
RL encodes a serine acetyltransferase."
RL J. Bacteriol. 173:5457-5469(1991).
CC
CC -i- FUNCTION: PROBABLE SERINE ACETYLTRANSFERASE REQUIRED FOR
CC OPTIMIZING THE EXPRESSION OF NITROGENASE ACTIVITY. NIFP MAY BE
CC REQUIRED TO BOOST RATES OF SYNTHESIS OR INTRACELLULAR
CC CONCENTRATIONS OF CYSTEINE OR METHIONINE.
CC
CC -i- CATALYTIC ACTIVITY: Acetyl-CoA + L-serine = CoA + O-acetyl-L-
CC serine.
CC
CC -i- SIMILARITY: BELONGS TO THE CYSE/LACA/LPYA/NOXL FAMILY OF
CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M60090; AAA22162.1; -.
CC PIR: DA3706; DA3706.
CC InterPro: IPR001451; Hexapep_transf.
CC Pfam: PF00132; hexapep; 4.
CC TIGRfams: TIGR01172; cyse; 1.
CC PROSITE: PS00101; HEXAPEP_TRANSFERASES; 1.
CC Nitrogen fixation; Transferase; Acyltransferase; Repeat;
CC Cysteine biosynthesis.
CC
CC SEQUENCE 269 AA; 28578 MW; E0BCC982B66FDB CRC64;
SQ

```

Query Match 8.9%; Score 7; DB 1; Length 269;  
Best Local Similarity 100.0%; Pred. No. 9.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 6 LPRPAV 12
Db 52 LPRPAV 58

```

RESULT 5

```

ID YMH7_CAEEL STANDARD: PRT; 380 AA.
AC P34474;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 31.8 kDa protein F58A4.7 in chromosome III.
GN F58A4.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Chordata; Rhabdillida; Rhabdilloidea;
OC Rhabdillidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN
RP SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Lathrop P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Spratt J.,
RA Woldman P.;

```

```

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC
CC -i- SUBCELLULAR LOCATION: Nuclear (Potential).
CC
CC -i- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. STRONGEST, TO TRANSCRIPTION FACTOR AP-4.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 222179; CA860167.1; -.
CC PIR: S40979; S40979.
CC WormPep: F58A4.7; CB01018.
CC InterPro: IPR001092; HLH_basic.
CC Pfam: PF00010; HLH; 1.
CC SMART: SM00353; HLH; 1.
CC PROSITE: PS00038; HLH_1; 1.
CC PROSITE: PS00888; HLH_2; 1.
CC Hypothetical protein; Transcription regulation; DNA-binding;
CC Nuclear protein.
CC
CC FT DNA_BIND 61 73 BASIC DOMAIN.
FT DOMAIN 74 113 BASIC DOMAIN.
FT SEQUENCE 380 AA; 40519 MW; EE2B52D1EB3EFD3 CRC64;
SQ

```

Query Match 8.9%; Score 7; DB 1; Length 380;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 46 APLSPS 52
Db 48 APLSPS 54

```

RESULT 6

```

ID CPML_ONCMY STANDARD: PRT; 499 AA.
AC Q92088;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 2M1 (EC 1.14.14.1) (CYPM1) (Lauric acid omega-6-
DE hydroxylase) (LMC1).
GN CYP2M1.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN
RP SEQUENCE FROM N.A.
RP STRAIN=Shasta; TISSUE=Liver;
RX MEDLINE=98248987; PubMed=9587416;
RA Yang Y.-H., Wang J.-L., Miranda C.L., Buhler D.R.;
RA "CYP2M1: Cloning, sequencing, and expression of a new cytochrome P450
RA from rainbow trout liver with fatty acid (omega-6)-hydroxylation
RA activity."
RL Arch. Biochem. Biophys. 352:271-280(1998).
RN
RP CHARACTERIZATION.
RP MEDLINE=90386619; PubMed=2403346;
RA Miranda C.L., Wang J.L., Henderson M.C., Williams D.E., Buhler D.R.;
RA "Regiospecificity in the hydroxylation of lauric acid by rainbow trout
RA hepatic cytochrome P450 isozyms."
RL Biochem. Biophys. Res. Commun. 171:537-542(1990).
CC
CC -i- FUNCTION: HAS (OMEGA-6)-HYDROXYLATION ACTIVITY TOWARD LAURIC ACID.
CC -i- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC
CC -i- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By

```

```

CC      similarity).
CC      -1- TISSUE SPECIFICITY: IN KIDNEY AND IN LIVER FROM JUVENILE AND
CC      SEXUALLY MATURE TROUT FROM BOTH SEXES.
CC      -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR EMBL: U16657; AAA62499.1;
CC      DR HSSP: P00179; IDR6.
CC      DR InterPro: IPR001128; Cytochrome_P450.
CC      DR Pfam: PF00067; P450.1.
CC      DR PRINTS: PR00385; P450.
CC      DR PROSITE: PS00086; CYTOCHROME_P450.1.
CC      KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
CC      Microsome; Endoplasmic reticulum.
CC      FT BINDING 441 441 HEME (BY SIMILARITY).
CC      SEQUENCE 499 AA; 56773 MW; 2D5C106C192ECBA2 CRC64;

Oy      48 LSPTSVP 54
Db      367 LSPTSVP 373

Query Match      8.9%; Score 7; DB 1; Length 499;
Best local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
ZEST_DROME      STANDARD;      PRT;      574 AA.
ID      ZEST_DROME
AC      P09956; Q24596; Q9V3F1;
DT      01-MAR-1989 (Rel. 10, Created)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Regulatory protein zeste.
GN      Z OR EG:BACH59J11.3 OR CG7803.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC      Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=88142560; PubMed=3125410;
RA      Mansukhani A., Gunaratne P.H., Sherwood P.W., Sneath B.J.,
RA      Goldberg M.L.;
RT      "Nucleotide sequence and structural analysis of the zeste locus of
RT      Drosophila melanogaster.";
RL      Mol. Gen. Genet. 211:121-128(1988).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      STRAIN-Oregon-R;
RX      MEDLINE=87218538; PubMed=3582372;
RA      Piroetta V., Manet E., Haddon E., Bickel S.E., Benson M.;
RT      "Structure and sequence of the Drosophila zeste gene.";
RL      EMBO J. 6:791-799(1987).
RN      [3]
RP      REVISIONS, AND SELF-ASSOCIATION.
RX      MEDLINE=90361011; PubMed=2118108;
RA      Bickel S.E., Piroetta V.;
RT      "Self-association of the Drosophila zeste protein is responsible for
RT      transvection effects.";
RL      EMBO J. 9:2959-2967(1990).
RN      [4]
RP      SEQUENCE FROM N.A.
RX      STRAIN-Berkeley;
RX      MEDLINE=20196006; PubMed=10731132;

```

```

RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
RA      Botkova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
RA      Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Foster C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA      Palazzolo M., Peltman G.S., Pan S., Pollard D., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,
RA      Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhu D., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195(2000).
RN      [5]
RP      SEQUENCE FROM N.A.
RX      STRAIN-Oregon-R;
RX      MEDLINE=20196011; PubMed=10731137;
RA      Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA      Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
RA      Driano S., Gloux S., Lelaune V., Mottier S., Galibert F., Botkova D.,
RA      Miñana B., Kafatos F.C., Louis C., Siden-Klimos I., Bolshakov S.,
RA      Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA      Modolell J., Peter A., Schoetcler P., Werner M., Moutkioti F.,
RA      Beinert N., Dowe G., Schaefer U., Jaecle H., Bucheton A.,
RA      Callister D.M., Campbell L.A., Darlamsitou A., Henderson N.S.,
RA      McMillan P.J., Sallies C., Tait E.A., Valentini P., Saunders R.D.C.,
RA      Glover D.M.;
RT      "From sequence to chromosome: the tip of the X chromosome of D.
RT      melanogaster.";
RL      Science 287:2220-2222(2000).
RN      [6]
RP      SELF-ASSOCIATION.
RX      MEDLINE=93259149; PubMed=8491197;
RA      Chen J.D., Piroetta V.;
RT      "Multimerization of the Drosophila zeste protein is required for
RT      efficient DNA binding.";
RL      EMBO J. 12:2075-2083(1993).
RN      [7]
RP      SEQUENCE OF 56-323 FROM N.A.
RX      STRAIN-ME-K1, ME-K2, ME-L11, ME-L12, ME-NJ1, and ME-NJ2;
RX      MEDLINE=93360802; PubMed=835601;
RA      Hey J., Kilman R.M.;
RT      "Population genetics and phylogenetics of DNA sequence variation at
RT      multiple loci within the Drosophila melanogaster species complex.";
RL      Mol. Biol. Evol. 10:804-822(1993).
CC      -1- FUNCTION: INVOLVED IN TRANSVECTION PHENOMENA (= SYNAPSIS-DEPENDENT
CC      GENE EXPRESSION), WHERE THE SYNAPTIC PAIRING OF CHROMOSOMES
CC      CARRYING GENES WITH WHICH ZESTE INTERACTS INFLUENCES THE

```

```
CC      EXPRESSION OF THESE GENES. ZESTIE BINDS TO DNA AND STIMULATES
CC      TRANSCRIPTION FROM A NEARBY PROMOTER.
CC      -1- SUBUNIT: SELF-ASSOCIATE FORMING COMPLEXES OF SEVERAL HUNDRED
CC      MONOMERS.
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: Y00049; CAA68262.1; ALT_SEQ.
CC      EMBL: X06743; CAA29918.1; ALT_SEQ.
CC      EMBL: L13043; AAA29026.1; -.
CC      EMBL: L13044; AAA29027.1; -.
CC      EMBL: L13045; AAA29028.1; -.
CC      EMBL: L13046; AAA29029.1; -.
CC      EMBL: L13047; AAA29030.1; -.
CC      EMBL: L13048; AAA29031.1; -.
CC      EMBL: AE003424; AAF45783.1; -.
CC      EMBL: AL133505; CAB63525.1; -.
CC      PIR: A26639; A26639.
CC      PIR: S01272; S01272.
CC      PIR: S12569; S12569.
CC      TRANSFAC: T00918; -.
CC      FLYBASE: FBgn0004050; Z.
CC      DNA-binding; Transcription regulation; Nuclear protein; Polymorphism.
CC      FT INIT_MER 0 0
CC      FT DOMAIN 1 46 HYDROPHOBIC.
CC      FT DNA_BIND 47 127 SPECIFIC, WITH ZESTIE LOCUS.
CC      FT DOMAIN 5 42 GLY-RICH.
CC      FT DOMAIN 152 430 GLN/ALA-RICH (OPR-REPAT INVOLVED IN
CC      TRANSCRIPTIONAL ACTIVATION OR REPRESSION
CC      AT DIFFERENT TARGET LOC1) (POTENTIAL).
CC      FT VARIANT 232 233 MISSING (IN STRAIN ME-K2).
CC      FT CONFLICT 496 496 S -> A (IN REF. 4 AND 5).
CC      FT SEQUENCE 574 AA; 61845 MW; 45DCD36CF72F5CF3 CRC64;
CC      -----
CC      Query Match 8.9%; Score 7; DB 1; Length 574;
CC      Best Local Similarity 100.0%; Pred. No. 18;
CC      Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC      -----
CC      QY 9 PAAVPP 15
CC      DB 449 PAAVPP 455
CC      -----
CC      RESULT 8
CC      SL56_HUMAN STANDARD; PRT; 635 AA.
CC      AC 091289;
CC      DT 16-OCT-2001 (Rel. 40; Created)
CC      DT 16-OCT-2001 (Rel. 40; Last sequence update)
CC      DT 16-OCT-2001 (Rel. 40; Last annotation update)
CC      DE Sodium-dependent multivitamin transporter (Na(+)-dependent
CC      DE multivitamin transporter).
CC      GN SLC5A6 OR SMT.
CC      OS Homo sapiens (Human).
CC      OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC      OC Mammalia; Eutheria; Primates; Carnivora; Homidae; Homo.
CC      OX NCBI_TaxID=9606;
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RC TISSUE=Intestine;
CC      RX MEDLINE=99268779; PubMed=10334869;
CC      RA Prasad P.D., Wang H., Huang W., Fei Y.-J., Leibach F.H., Devoe L.D.,
CC      RA Ganapathy V.;
CC      RA "Molecular and functional characterization of the intestinal Na+-
CC      RA dependent multivitamin transporter.";
CC      RA Arch. Biochem. Biophys. 366:95-106(1999).
CC      CC -1- FUNCTION: TRANSPORTS PANTOTHENATE, BIOTIN AND LIPONATE IN THE
CC      CC PRESENCE OF SODIUM.
CC      CC -1- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC      CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AF116241; AAD37502.1; -.
CC      EMBL: AF069307; AAD31727.1; -.
CC      EMBL: AF081571; AAD37481.1; -.
CC      GeneW: HGNC:11041; SLC5A6.
CC      DR MIM: 604024; -.
CC      DR InterPro: IPR001734; Na/solut_symport.
CC      DR Pfam: PF00474; SSF: 1.
CC      DR TIGRfams: TIGR00813; sss: 1.
CC      DR PROSITE: PS00456; NA_SOLUT_SYM_1; 1.
CC      DR PROSITE: PS00457; NA_SOLUT_SYM_2; FALSE_NEG.
CC      DR PROSITE: PS0283; NA_SOLUT_SYM_3; 1.
CC      KW Transport; Transmembrane; Sodium transporter; Symport; Glycoprotein.
CC      FT TRANSMEM 24 44 POTENTIAL.
CC      FT TRANSMEM 68 88 POTENTIAL.
CC      FT TRANSMEM 101 121 POTENTIAL.
CC      FT TRANSMEM 143 163 POTENTIAL.
CC      FT TRANSMEM 176 196 POTENTIAL.
CC      FT TRANSMEM 199 219 POTENTIAL.
CC      FT TRANSMEM 256 276 POTENTIAL.
CC      FT TRANSMEM 297 317 POTENTIAL.
CC      FT TRANSMEM 336 356 POTENTIAL.
CC      FT TRANSMEM 396 416 POTENTIAL.
CC      FT TRANSMEM 428 448 POTENTIAL.
CC      FT TRANSMEM 456 476 POTENTIAL.
CC      FT TRANSMEM 528 548 POTENTIAL.
CC      FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT SEQUENCE 635 AA; 68701 MW; 4F10369A3916564F CRC64;
CC      -----
CC      Query Match 8.9%; Score 7; DB 1; Length 635;
CC      Best Local Similarity 100.0%; Pred. No. 19;
CC      Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC      -----
CC      QY 46 APLSPS 52
CC      DB 9 APLSPS 15
CC      -----
CC      RESULT 9
CC      GCP3_MOUSE STANDARD; PRT; 677 AA.
CC      AC P38854;
CC      DT 15-JUN-2002 (Rel. 41; Created)
CC      DT 15-JUN-2002 (Rel. 41; Last sequence update)
CC      DT 15-JUN-2002 (Rel. 41; Last annotation update)
CC      DE Gamma-tubulin complex component 3 (GCP-3) (Fragment).
CC      GN GCP3.
CC      OS Mus musculus (Mouse).
CC      OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC      OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=99268779; PubMed=10334869;
RA Prasad P.D., Wang H., Huang W., Fei Y.-J., Leibach F.H., Devoe L.D.,
RA Ganapathy V.;
RA "Molecular and functional characterization of the intestinal Na+-
RA dependent multivitamin transporter.";
RA Arch. Biochem. Biophys. 366:95-106(1999).
CC -1- FUNCTION: TRANSPORTS PANTOTHENATE, BIOTIN AND LIPONATE IN THE
CC PRESENCE OF SODIUM.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF116241; AAD37502.1; -.
CC EMBL: AF069307; AAD31727.1; -.
CC EMBL: AF081571; AAD37481.1; -.
CC GeneW: HGNC:11041; SLC5A6.
CC DR MIM: 604024; -.
CC DR InterPro: IPR001734; Na/solut_symport.
CC DR Pfam: PF00474; SSF: 1.
CC DR TIGRfams: TIGR00813; sss: 1.
CC DR PROSITE: PS00456; NA_SOLUT_SYM_1; 1.
CC DR PROSITE: PS00457; NA_SOLUT_SYM_2; FALSE_NEG.
CC DR PROSITE: PS0283; NA_SOLUT_SYM_3; 1.
CC KW Transport; Transmembrane; Sodium transporter; Symport; Glycoprotein.
CC FT TRANSMEM 24 44 POTENTIAL.
CC FT TRANSMEM 68 88 POTENTIAL.
CC FT TRANSMEM 101 121 POTENTIAL.
CC FT TRANSMEM 143 163 POTENTIAL.
CC FT TRANSMEM 176 196 POTENTIAL.
CC FT TRANSMEM 199 219 POTENTIAL.
CC FT TRANSMEM 256 276 POTENTIAL.
CC FT TRANSMEM 297 317 POTENTIAL.
CC FT TRANSMEM 336 356 POTENTIAL.
CC FT TRANSMEM 396 416 POTENTIAL.
CC FT TRANSMEM 428 448 POTENTIAL.
CC FT TRANSMEM 456 476 POTENTIAL.
CC FT TRANSMEM 528 548 POTENTIAL.
CC FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT SEQUENCE 635 AA; 68701 MW; 4F10369A3916564F CRC64;
CC -----
CC Query Match 8.9%; Score 7; DB 1; Length 635;
CC Best Local Similarity 100.0%; Pred. No. 19;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 46 APLSPS 52
CC DB 9 APLSPS 15
CC -----
CC RESULT 9
CC GCP3_MOUSE STANDARD; PRT; 677 AA.
CC AC P38854;
CC DT 15-JUN-2002 (Rel. 41; Created)
CC DT 15-JUN-2002 (Rel. 41; Last sequence update)
CC DT 15-JUN-2002 (Rel. 41; Last annotation update)
CC DE Gamma-tubulin complex component 3 (GCP-3) (Fragment).
CC GN GCP3.
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-Breast tumor;		
RA	Strausberg R.;		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: Gamma-tubulin complex is necessary for microtubule nucleation at the centrosome (By similarity).		
CC	-1- SUBUNIT: Gamma-tubulin complex is composed of gamma-tubulin, GCP2, GCP3, GCP4, GCP5 and GCP6 (By similarity).		
CC	-1- SUBCELLULAR LOCATION: Centrosome (By similarity).		
CC	-1- SIMILARITY: BELONGS TO THE GCP FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/isb-sib.ch">http://www.isb-sib.ch/announce/isb-sib.ch</a> ).		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; BC025647; AAH25647.1; -		
KW	Microtubules.		
FT	NON_TER 1 1		
FT	DOMAIN 595 599 POLY-GLU		
SO	SEQUENCE 677.AA: 78348 MW: 164854473575E33C4 CRC64:		

Query Match	8.9%;	Score 7;	DB 1;	Length 677;
Best Local Similarity	100.0%;	Pred. No. 21;		
Matches	7;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

Qy	24	HPVLSFL	30
Db	194	HPVLSFL	200

RESULT 10	
CAD2_MOUSE	
ID	STANDARD:
CAD2_MOUSE	PRT: 906 AA.

DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neural-cadherin precursor (N-cadherin) (Cadherin-2).  
 CN CDM2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;

RP SEQUENCE FROM N.A.  
RX MEDLINE=89346748; PubMed=2762814;  
RA Miyetani S., Shimamura K., Hatta M., Nagafuchi A., Nose A.,  
RA Matsunaga M., Hatta K., Takeichi M.,  
RT "Neural cadherin: role in selective cell-cell adhesion.",  
RL Science 245:631-635(1989).

RP SEQUENCE FROM N.A.  
RA Tamura K.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP PARTIAL SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RX MEDLINE=92409332; PubMed=1528849;  
RA Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A., Takeichi M.;  
RT "Genomic structure and chromosomal mapping of the mouse N-cadherin  
RL gene".  
Proc. Natl. Acad. Sci. U.S.A. 89:8443-8447(1992).

RP DEVELOPMENTAL STAGE.  
RC STRAIN=C57BL/6; TISSUE=Testis;  
RX MEDLINE=97033837; PubMed=8679495;  
RA Munro S.B., Blaschuk O.W.;

RA	"A comprehensive survey of the cadherins expressed in the testes of fetal, immature, and adult mice utilizing the polymerase chain reaction.";		
RT	Biol. Reprod. 55:822-827(1996).		
RL	[5]		
RN	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 160-267.		
RP	MEDLINE=95191680; PubMed=7885471;		
RX	Shrebel L., Fannon A.M., Kwong P.D., Thompson A., Lehmann M.S.,		
RA	Grubel G., Legrand J.-F., Als-Nielsen J., Colman D.R.,		
RA	Hendrickson W.A.;		
RT	"Structural basis of cell-cell adhesion by cadherins.";		
RL	Nature 374:327-337(1995).		
RN	[6]		
RP	X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS) OF 160-374.		
RX	MEDLINE=998318235; PubMed=9655503;		
RA	Tamura K., Shan W.S., Hendrickson W.A., Colman D.R., Shapiro L.;		
RT	"Structure-function analysis of cell adhesion by neural (N-) cadherin.";		
RL	Neuron 20:1153-1163(1998).		
CC	-1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS. THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN NEURONAL RECOGNITION MECHANISM.		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED AT ALL STAGES OF TESTICULAR DEVELOPMENT WITH HIGHEST LEVELS FOUND IN TESTES OF 21-DAY-OLD MICE.		
CC	-1- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-slb.ch/announce/">http://www.isb-slb.ch/announce/</a> or send an email to <a href="mailto:license@isb-slb.ch">license@isb-slb.ch</a> ).		
CC	-----		
DR	EMBL; M31131; AAA37933.1; -		
DR	EMBL; AB008811; BAA3549.1; -		
DR	EMBL; S45011; AAB23356.1; -		
DR	PIR; A32759; IJMSCN.		
DR	PDB; 1NCG; 10-JUL-95.		
DR	PDB; 1NCH; 10-JUL-95.		
DR	PDB; 1NCI; 10-JUL-95.		
DR	PDB; 1NCJ; 18-MAR-99.		
DR	MCD; MGI:88355; Cdh2.		
DR	InterPro: IPR002126; Cadherin.		
DR	InterPro: IPR000233; Cadherin_C-term.		
DR	Pfam: PF00028; cadherin; 5.		
DR	Pfam: PF01049; Cadherin_C-term; 1.		
DR	PRINTS; PRO0205; CADHERIN.		
DR	SMART; SM00112; CA; 5.		
DR	PROSITE; PS00232; CADHERIN_1; 3.		
DR	PROSITE; PS0268; CADHERIN_2; 5.		
KW	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat; Signal; 3D-structure		
KW	Signal, 3D-structure		
FT	1	23	POTENTIAL.
FT	PROPEP	159	POTENTIAL.
FT	CHAIN	160	NEURAL-CADHERIN.
FT	DOMAIN	160	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	725	POTENTIAL.
FT	DOMAIN	747	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	160	CADHERIN 1.
FT	DOMAIN	268	CADHERIN 2.
FT	DOMAIN	383	CADHERIN 3.
FT	DOMAIN	498	CADHERIN 4.
FT	DOMAIN	604	CADHERIN 5.
FT	DOMAIN	863	SER-RICH.
FT	CARBOHYD	190	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	273	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	325	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	402	N-LINKED (GLCNAC. . .) (POTENTIAL).

```
CC MICE
CC -1 SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on lists
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M31131; AAA37553.1; -.
CC DR EMBL; AB008911; BAA23549.1; -.
CC DR EMBL; S45011; AAB23356.1; -.
CC DR PIR; A32759; IJMSCN.
CC DR PDB; INCG; 10-JUL-95.
CC DR PDB; INCJ; 10-JUL-95.
CC DR PDB; INCI; 10-JUL-95.
CC DR PDB; INCT; 18-MAR-99.
CC DR MCP; MGI; 88355; CdH2.
CC DR InterPro; IPR002126; Cadherin.
CC DR InterPro; IPR000233; Cadherin_C_term.
CC DR Pfam; PF00028; cadherin; 5.
CC DR Pfam; PF01049; Cadherin_C_term; 1.
CC DR PRINTS; PRO0205; CADHERIN.
CC DR SMART; SM00112; CA; 5.
CC DR PROSITE; PS00232; CADHERIN_1; 3.
CC DR PROSITE; PS50268; CADHERIN_2; 5.
CC KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC Signal; 3D-structure.
CC FT SIGNAL; 1 23 POTENTIAL.
CC FT PROPEP; 24 159 POTENTIAL.
CC FT CHAIN; 160 906 NEURAL-CADHERIN.
CC FT DOMAIN; 160 724 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM; 725 746 POTENTIAL.
CC FT DOMAIN; 747 906 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN; 160 267 CADHERIN 1.
CC FT DOMAIN; 268 382 CADHERIN 2.
CC FT DOMAIN; 383 497 CADHERIN 3.
CC FT DOMAIN; 498 603 CADHERIN 4.
CC FT DOMAIN; 604 717 CADHERIN 5.
CC FT DOMAIN; 863 878 SER-RICH.
CC FT CARBOHYD; 190 190 N-LINKED (GLCNAc; . .) (POTENTIAL).
CC FT CARBOHYD; 273 273 N-LINKED (GLCNAc; . .) (POTENTIAL).
CC FT CARBOHYD; 325 325 N-LINKED (GLCNAc; . .) (POTENTIAL).
CC FT CARBOHYD; 402 402 N-LINKED (GLCNAc; . .) (POTENTIAL).
```

FT CARBOHYD	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD	273	273	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD	325	325	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD	402	402	N-LINKED (GLCNAC. . .)	(POTENTIAL).

```
FT CARBOHYD 572 572 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 692 692 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 7 9 GRG -> APR (IN REF. 2).
FT CONFLICT 565 565 Y -> N (IN REF. 2).
FT CONFLICT 567 567 Q -> K (IN REF. 2).
FT CONFLICT 624 624 A -> T (IN REF. 2).
SQ SEQUENCE 906 AA; 99761 MW; 0206741B71707B7 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 VAVNLSR 76
Db 117 VAVNLSR 123

RESULT 11
GCP3_HUMAN STANDARD; PRT; 907 AA.
ID GCP3_HUMAN 043631; 060852; 060853; 096179;
AC 096CW5: 043631; 060852; 060853; 096179;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gamma-tubulin complex component 3 (GCP-3) (Spindle pole body protein
GN GCP3
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Cervical carcinoma; PubMed=956967;
RA MEDLINE=98234402; PubMed=956967;
RT Murphy S.M., Urban L., Stearns T.;
RT "The mammalian gamma-tubulin complex contains homologues of the yeast
RL spindle pole body components spc97p and spc98p.";
RL J. Cell Biol. 141:663-674(1998).
[2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND CHARACTERIZATION.
RA MEDLINE=98234404; PubMed=956969;
RT Tassin A.-M., Celati C., Moudjou M., Bornens M.;
RT "Characterization of the human homologue of the yeast spc98p and its
RL association with gamma-tubulin.";
RL J. Cell Biol. 141:689-701(1998).
[3]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RC TISSUE=Eye, and Muscle;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Gamma-tubulin complex is necessary for microtubule
CC nucleation at the centrosome.
CC -1- SUBUNIT: Gamma-tubulin complex is composed of gamma-tubulin, GCP2,
CC GCP3, GCP4, GCP5 and GCP6.
CC -1- SUBCELLULAR LOCATION: Centrosome.
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; may be
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -1- SIMILARITY: BELONGS TO THE GCP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF043378; AAC39727.1; -
DR EMBL: AU003061; CAA05833.1; -
DR EMBL: AJ003062; CAA05833.1; -
```

```
DR EMBL: BC013781; AAH13781.1; -
DR EMBL: BC007763; AAH07763.1; -
KW Microtubules; Alternative splicing.
FT DOMAIN 825 829 POLY-GLU
FT VARSPPLIC 817 824 GQGVATAA -> VEMKLYCV (IN ISOFORM 2).
FT VARSPPLIC 825 907 MISSING (IN ISOFORM 2).
FT VARSPPLIC 391 434 RKGELASAVHATKTGDPMKSLVQHILSLVSHPEVSLY
FT VARSPPLIC RMT -> PPRVPTHTVPTPRDPTPRDPTPRHVPPTPRV
FT VARSPPLIC WHISCPRTLL (IN ISOFORM 3).
FT VARSPPLIC 435 907 MISSING (IN ISOFORM 3).
FT CONFLICT 208 208 T -> S (IN REF. 2).
FT CONFLICT 361 361 S -> I (IN REF. 2).
SQ SEQUENCE 907 AA; 103570 MW; 70FE2FDB7C80344D CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 HPVLSFL 30
Db 424 HPVLSFL 430

RESULT 12
ST10_HUMAN STANDARD; PRT; 968 AA.
ID ST10_HUMAN 09UW4;
AC 094804; 09UW4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine-protein kinase 10 (EC 2.7.1.37) (Lymphocyte-oriented
GN STK10 OR LOK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99216434; PubMed=10199912;
RA Kuramochi S., Matsuda Y., Okamoto M., Kitamura F., Yonekawa H.,
RA Kurasuyama H.;
RT "Molecular cloning of the human gene STK10 encoding lymphocyte-
RT oriented kinase, and comparative chromosomal mapping of the human,
RT mouse, and rat homologues.";
RL Immunogenetics 49:369-375(1999).
[2]
RP SEQUENCE OF 814-968 FROM N.A.
RC TISSUE=Testis;
RA Bloeker H., Boecker M., Brandt P., Mewes H.-W., Gassenhuber J.,
RA Wiemann S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CAN ACT ON SUBSTRATES SUCH AS MYELIN BASIC PROTEIN AND
CC HISTONE IIA ON SERINE AND THREONINE RESIDUES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN LYMPHOID ORGANS.
CC -1- PTM: AUTOPHOSPHORYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB015718; BAA35073.1; -
DR EMBL: AL133081; CAB61400.1; -
DR HSSP: P24941; IHCL.
DR Genew: HGNC:11388; STK10.
DR MIM: 603919; -
```

DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00220; S\_TKC; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transfaser: Serine/threonine-protein kinase: ATP-binding;  
 KW Phosphorylation; Coiled coil.  
 KW DOMAIN 36  
 FT DOMAIN 36  
 FT DOMAIN 573  
 FT NP\_BIND 750  
 FT NP\_BIND 42  
 FT BINDING 65  
 FT ACT\_SITE 157  
 FT ACT\_SITE 157  
 SQ SEQUENCE 968 AA; 112134 MW; 15E245193EC553D CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AVVPPLR 17  
 Db 405 AVVPPLR 411

RESULT 13  
 PSBK\_PINTH STANDARD; PRT; 56 AA.

AC P41598;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Photosystem II reaction center protein K precursor (PSII-K).  
 GN PSBK.  
 OS Pinus thunbergii (Green pine) (Japanese black pine).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferales; Pinaceae; Pinus.  
 OX NCBI\_TaxID=3350;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95024047; PubMed=7937893;  
 RA Waksugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,  
 RA Sugitara M.;  
 RT "Loss of all ndh genes as determined by sequencing the entire  
 RT chloroplast genome of the black pine Pinus thunbergii.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).  
 CC -! FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER  
 CC OF PHOTOSYSTEM II.  
 CC -! SIMILARITY: BELONGS TO THE PSBK FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: D17510; BAA04312.1; ALT\_INIT.  
 DR InterPro: IPR003687; PSII\_PSBK.  
 DR Pfam: PF02533; PSBK; 1.

KW Photosystem II; Chloroplast.  
 FT PROPEP 1  
 FT CHAIN 20 56  
 SQ SEQUENCE 56 AA; 6346 MW; 18BIDEFF198ACACA CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Score 6; DB 1; Length 56;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 25 PVLSFL 30  
 Db 39 PVLSFL 44

RESULT 14  
 YCXS\_OENHO STANDARD; PRT; 111 AA.  
 AC Q9WNT3;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 12.8 kDa protein in ycf9-trns intergenic region (ORF111).  
 OS Oenothera hookeri (Hooker's evening primrose).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosida II; Myrtales; Onagraceae; Oenothera.  
 OX NCBI\_TaxID=85636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Johansen;  
 RX MEDLINE=20309318; PubMed=10852478;  
 RA Hupler H., Swiatek M., Hornung S., Herrmann R.G., Maier R.M.,  
 RA Chiu W.-T., Sears B.;  
 RT "Complete nucleotide sequence of the Oenothera elata plastid  
 RT chromosome, representing plastome I of the five distinguishable  
 RT Oenothera plastomes.";  
 RL Mol. Gen. Genet. 263:581-585(2000).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: AJ271079; CAB67142.1; -  
 KW Chloroplast; Hypothetical protein.  
 SQ SEQUENCE 111 AA; 12814 MW; E5E0CE989317F140 CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 72 VNLSRY 77  
 Db 100 VNLSRY 105

RESULT 15  
 MF14\_MAIZE STANDARD; PRT; 126 AA.  
 AC 001900;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE MF14 protein precursor.  
 GN MF14.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC Clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. B10; TISSUP=Tassel;  
 RX MEDLINE=94004987; PubMed=8401606;  
 RX Wright S.Y., Suer W.-M., Bell P.J., Vaundin M., Greenland A.J.;  
 RT "Isolation and characterization of male flower cDNAs from maize.";

```

RL Plant J. 3:41-49(1993).
CC -!- TISSUE SPECIFICITY: ENHANCED EXPRESSION IN MALE FLOWERS.
CC ACCUMULATES IN THE TAPETUM.
CC -!- DEVELOPMENTAL STAGE: ASSOCIATED WITH MICROSPOROGENESIS AND
CC DECLINES AS MATURE POLLEN IS PRODUCED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X67323; CAA4737.1; -.
DR PIR: S25104; S25104.
DR MaizeDB: 69183; -.
KW Signal.
FT SIGNAL 1 23 OR 24, OR 26 (POTENTIAL).
FT CHAIN 24 126 MFS14 PROTEIN.
SQ SEQUENCE 126 AA; 12653 MW; 67F2813AF8C555 CRC64;

Query Match 7.68; Score 6; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RPAVP 13
   |||||
Db 38 RPAVP 43

```

Search completed: March 4, 2003, 12:45:36  
 Job time : 7.39357 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:42:02 ; Search time 8.7249 seconds  
(without alignments)  
1865.663 Million cell updates/sec

Title: US-09-234-208B-1

Perfect score: 79  
Sequence: 1 GTHSLPPRAAIVPPLRMQP.....VGRGPPDAHVAVNLRYEG 79

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	83.5	419	4 Q9UK79	Q9UK79 homo sapien
2	8	10.1	217	5 Q9M6W6	Q9M6W6 drosophila
3	8	10.1	481	10 Q945M4	Q945M4 arabidopsis
4	8	10.1	520	16 Q66518	Q66518 aquilex aeo
5	8	10.1	865	10 Q22774	Q22774 arabidopsis
6	8	10.1	879	10 Q41010	Q41010 pisum sativ
7	8	10.1	1469	10 Q9LKR1	Q9LKR1 pisum sativ
8	8	10.1	1503	10 Q81283	Q81283 arabidopsis
9	7	8.9	27	4 Q9HD19	Q9HD19 homo sapien
10	7	8.9	39	5 Q8WQJ2	Q8WQJ2 plasmodium
11	7	8.9	121	16 Q8YZW8	Q8YZW8 anabaena sp
12	7	8.9	146	15 Q8UJ38	Q8UJ38 human immun
13	7	8.9	146	15 Q8UJ36	Q8UJ36 human immun
14	7	8.9	178	10 Q9FMC7	Q9FMC7 arabidopsis
15	7	8.9	200	5 Q95PE1	Q95PE1 amblyomma a
16	7	8.9	241	17 Q8TW69	Q8TW69 methanopyru

17	7	8.9	245	4 Q14596	Q14596 homo sapien
18	7	8.9	268	2 Q9EWM9	Q9EWM9 streptomyce
19	7	8.9	328	10 Q81138	Q81138 arabidopsis
20	7	8.9	350	13 Q91AL2	Q91AL2 gallus gall
21	7	8.9	382	5 Q9U3E1	Q9U3E1 caenorhabdi
22	7	8.9	389	16 Q8UEF8	Q8UEF8 agrobacteri
23	7	8.9	392	12 Q9Q940	Q9Q940 shope fibro
24	7	8.9	397	16 Q913T2	Q913T2 pseudomonas
25	7	8.9	416	16 Q9KZK5	Q9KZK5 streptomyce
26	7	8.9	436	16 Q67814	Q67814 aquilex aeo
27	7	8.9	445	16 Q8R814	Q8R814 thermococci
28	7	8.9	464	16 Q8YST7	Q8YST7 anabaena sp
29	7	8.9	469	5 Q8WQ63	Q8WQ63 drosophila
30	7	8.9	469	5 Q8WPA9	Q8WPA9 drosophila
31	7	8.9	469	5 Q8WPA5	Q8WPA5 drosophila
32	7	8.9	471	5 Q8WQ65	Q8WQ65 drosophila
33	7	8.9	471	5 Q8WQ64	Q8WQ64 drosophila
34	7	8.9	471	5 Q8WPC6	Q8WPC6 drosophila
35	7	8.9	471	5 Q8WPA8	Q8WPA8 drosophila
36	7	8.9	471	5 Q8WPA6	Q8WPA6 drosophila
37	7	8.9	473	5 Q8WQ66	Q8WQ66 drosophila
38	7	8.9	473	5 Q8WPA9	Q8WPA9 drosophila
39	7	8.9	475	5 Q8WQ67	Q8WQ67 drosophila
40	7	8.9	495	2 Q52799	Q52799 amycolatops
41	7	8.9	595	5 Q18900	Q18900 caenorhabdi
42	7	8.9	635	4 Q969Y5	Q969Y5 homo sapien
43	7	8.9	798	2 Q8YR14	Q8YR14 myxococcus
44	7	8.9	840	16 Q9X8B0	Q9X8B0 streptomyce
45	7	8.9	1009	15 Q901A3	Q901A3 human immun

#### ALIGNMENTS

RESULT 1	
ID Q9UK79	PRELIMINARY; PRT; 419 AA.
AC Q9UK79;	
DT 01-MAY-2000 (TREMBLrel. 13, Created)	
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE Herstatin.	
GN HER-2.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.	
OX NCBI_TaxID=9606;	
RP [1]	
RP SEQUENCE FROM N.A.	
RA MEDLINE-99415951; PubMed-10485918;	
RX Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.:	
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted	
RT autolnhibitor.";	
RT Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).	
RP [2]	
RP SEQUENCE FROM N.A.	
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Hennen W.D.:	
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.	
DR EMBL; AF17761; AAD56009.2; -	
DR InterPro: IPR000494; EGFR_L_domain.	
DR Pfam: PF00757; Furin-like.	
DR Pfam: PF01030; Recep_L_domain; 1.	
DR SMART; SM00261; FU; 1.	
SQ SEQUENCE 419 AA; 45472 MW; EC1BE347E2D030C CRC64;	
Query Match	
Best local similarity 83.5%; Score 66; DB 4; Length 419;	
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy 7 PRPAAVPVLRMQPGPAHVLISFLRPSWDLVSATFSLPLAPLSPTSPVSVGRGDP 66	
Db 347 PRPAAVPVLRMQPGPAHVLISFLRPSWDLVSATFSLPLAPLSPTSPVSVGRGDP 406	

OY 67 DAHAVAV 72  
 DB 407 DAHAVAV 412

## RESULT 2

O9N6M6 PRELIMINARY: PRT: 217 AA.  
 AC O9N6M6:  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE HLHgamma protein.  
 GN HLH-GAMMA OR E.  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7224;  
 GN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cooper M.T.D., Tyler D.M., Furriols M., Chalkiadaki A., Delidakis C.,  
 RA Bray S.;  
 RT "Spatially restricted factors co-operate with Notch in the regulation  
 RT of Enhancer of split genes.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ276313; CAB7019.1; -;  
 DR Flybase: FBgn013370; dhyd\HLH-gamma.  
 DR InterPro: IPR001092; HLH\_basic.  
 DR InterPro: IPR003650; Orange.  
 DR Pfam: PF00010; HLH; 1.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00511; ORANGE; 1.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; UNKNOWN.1.  
 SQ SEQUENCE 217 AA: 2408 MW: 3917C58F322B4ED CRC64;

Query Match 10.1%; Score 8; DB 5; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 SVPSVPS 59  
 DB 159 SVPSVPS 166

## RESULT 3

O945M4 PRELIMINARY: PRT: 481 AA.  
 AC O945M4:  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE AT4G02510/T10P11.19.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 GN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Natuska M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF412063; AAL06516.1; -;  
 DR TIGRfams: TIGR00993; 3a0901s04IAP86; 1.

SQ SEQUENCE 481 AA: 53603 MW: D3DD3F8AF6EE7E3A CRC64;

Query Match 10.1%; Score 8; DB 10; Length 481;  
 Best Local Similarity 100.0%; Pred. No. 7.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 PAAPVPL 16  
 DB 200 PAAPVPL 207

## RESULT 4

O66518 PRELIMINARY: PRT: 520 AA.  
 AC O66518:  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein AO\_116.  
 GN AO\_116.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 OX NCBI\_TaxID=63563;  
 GN [1]  
 RP SEQUENCE FROM N.A.  
 RA RC STRAIN-VF5;  
 RX MEDLINE-98196666; PubMed-9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek K., Sneed M.A., Keller M., Anjay M., Huber R.,  
 RA Felman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus.";  
 RL Nature 392:353-358(1998).  
 DR EMBL: AE000674; AAC06482.1; -;  
 DR InterPro: IPR001450; 4Fe4S\_FERREDOXIN.  
 DR PROSITE: PS00198; 4Fe4S\_FERREDOXIN; UNKNOWN.1.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 520 AA: 59472 MW: 979DB8B6BF2A20F1 CRC64;

Query Match 10.1%; Score 8; DB 16; Length 520;  
 Best Local Similarity 100.0%; Pred. No. 8.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 PLSPTSVP 54  
 DB 2 PLSPTSVP 9

## RESULT 5

O22774 PRELIMINARY: PRT: 865 AA.  
 AC O22774:  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative chloroplast outer envelope 86-kDa protein.  
 GN T10P11.19 OR AT4G02510.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 GN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CV. COLUMBIA;  
 RA Kaplan N., Johnson D., Schutz K., Gnoj L., Hoffman J., Till S.,  
 RA de la Bastide M., Granat S., Hameed A., Gottesman T., Hasegawa A.,  
 RA Shody N., Parnell L., Dedhia N., Johnson A.F., Lodhi M.,  
 RA Martensen R., Chen E.Y., Wilson R., McCombie W.R.;  
 RT "Sequence of A. thaliana BAC T10P11 from chromosome IV.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,  
RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,  
RA Shekher M., Schütz K., See L.H., Swaby I., Hebermann K., Dedhia N.N.,  
RA Mewes H.W., Lemcke K., Mayer K.F.X.,  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X.,  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC002330; AAC78265.2; -;  
DR EMBL: AL161494; CAB80744.1; -;  
DR TIGRFAWS: TIGR00993; 3a0901s04IAP86; 1;  
SQ SEQUENCE 865 AA; 94388 MW; 5CFD60339AAB39D CRC64;

Query Match 10.1%; Score 8; DB 10; Length 865;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 PAAVPVPL 16  
DB 584 PAAVPVPL 591

RESULT 6  
Q41010 PRELIMINARY; PRT; 879 AA.  
AC Q41010: Q41030;

DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Chloroplast outer envelope protein OEP86 precursor (GTP-binding  
protein IAP86).

DE IAP86.

OS Pisum sativum (Garden pea).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.

OX NCBI\_TaxID=3888;

RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=CV. MIRANDA; TISSUE=LEAF.

RX MEDLINE=95099324; PubMed=7801125;  
RA Hirsch S., Muckel E., Heemeyer F., von Heijne G., Soli J.,  
RT "A receptor component of the chloroplast protein translocation  
machinery";

RL Science 266:1989-1992(1994).

RN [2]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 551-561; 665-675 AND 867-879.

RX MEDLINE=95063938; PubMed=7973656;  
RA Kessler F., Blobel G., Patel H.V., Schnell D.J.,  
RT Identification of two GTP-binding proteins in the chloroplast protein  
import machinery";

RL Science 266:1035-1039(1994).

CC -1- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO CHLOROPLASTS.  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST OUTER MEMBRANE.

CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-31 IS THE INITIATOR.  
DR EMBL: Z31581; CAAB3453.1; -;

DR EMBL: L36857; AAA53276.1; -;

DR TIGRFAWS: TIGR00993; 3a0901s04IAP86; 1.  
KM Chloroplast; ATP-binding, outer membrane; GTP-binding.

FT TRANSIT 1 146 CHLOROPLAST.  
FT CHAIN 147 879 CHLOROPLAST OUTER ENVELOPE PROTEIN  
OEP86.

FT NP\_BIND 242 252 ATP (POTENTIAL).  
FT CONFLICT 373 375 GPS -> EOO (IN REF. 2)

SQ SEQUENCE 879 AA; 96548 MW; 3BC1C4E9743A0280 CRC64;

Query Match 10.1%; Score 8; DB 10; Length 879;

Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 PAAVPVPL 16  
DB 599 PAAVPVPL 606

RESULT 7  
Q9LKR1 PRELIMINARY; PRT; 1469 AA.  
AC Q9LKR1;

DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Chloroplast protein import component Toc159.

OS Pisum sativum (Garden pea).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.

OX NCBI\_TaxID=3888;

RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=20177862; PubMed=10712545;  
RA Chen K., Chen X., Schnell D.J.,  
RT "Initial binding of preproteins involving the toc159 receptor can be  
bypassed during protein import into chloroplasts.";

RL Plant Physiol. 122:813-822(2000).

DR EMBL: AF262939; AAF75761.1; -;

DR TIGRFAWS: TIGR00993; 3a0901s04IAP86; 1.  
SQ SEQUENCE 1469 AA; 158676 MW; AA18A7C3DB5746C4 CRC64;

Query Match 10.1%; Score 8; DB 10; Length 1469;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 PAAVPVPL 16  
DB 1189 PAAVPVPL 1196

RESULT 8  
O81283 PRELIMINARY; PRT; 1503 AA.  
AC O81283;

DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE T14P8.24 protein.

GN T14P8.24.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;  
RA Washu;

RT "The A. thaliana Genome Sequencing Project.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;  
RA Kalicki J., Elliott G., Cloud J.,  
RT "The sequence of A. thaliana T14P8.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [3]  
RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;  
RA Waterston R.;

RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF069298; AAC19285.1; -;

DR TIGR00993; 3a0901s04IAP86; 1.  
SQ SEQUENCE 1503 AA; 160818 MW; 04AEDE84C1BED3f6 CRC64;  
Query Match  
Best Local Similarity 100.0%; Score 8; DB 10; Length 1503;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 PAAVPVL 16  
Db 1222 PAAVPVL 1229  
RESULT 9  
Q9HD19 PRELIMINARY; PRT; 27 AA.  
AC Q9HD19;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Sodium-dependent multivitamin transporter (Fragment).  
GN SMVT OR SLC5A6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEART;  
RT Rubin S.A.; Dey S.; Chatterjee N.; Said H.M.;  
RT "Molecular characterization of the human heat SMVT cDNA.";  
RN Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21952373; PubMed-11955628;  
RA Dey S.; Subramanian V.S.; Chatterjee N.S.; Rubin S.A.; Said H.M.;  
RT "Characterization of the 5' regulatory region of the human sodium-  
dependent multivitamin transporter, hSMVT.";  
RL Biochim. Biophys. Acta 1574:187-192(2002).  
DR EMBL; AF288781; AAG00587.1; -;  
DR EMBL; AF442149; AAL84706.1; -;  
DR EMBL; AF442150; AAL84707.1; -;  
FT NON\_TER 27  
SQ SEQUENCE 27 AA; 2620 MW; 27CDCC338BD5E9D CRC64;  
Query Match  
Best Local Similarity 8.9%; Score 7; DB 4; Length 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 46 APLSPTS 52  
Db 9 APLSPTS 15  
RESULT 10  
Q8W0U2 PRELIMINARY; PRT; 39 AA.  
AC Q8W0U2;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Rhop-3 (Fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PF452;  
RA Wang T.; Sam-Yellowe T.Y.;  
RT "Conservation of the rhopty protein Rhop-3 among human and rodent  
RT Plasmodium species and its potential as a malaria vaccine candidate.";  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY044910; AAL65267.1; -;  
FT NON\_TER 1

FT NON\_TER 39  
SQ SEQUENCE 39 AA; 4498 MW; 853A12A3950B87A7 CRC64;  
Query Match  
Best Local Similarity 8.9%; Score 7; DB 5; Length 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 54 PISPVSV 60  
Db 15 PISPVSV 21  
RESULT 11  
Q8YZW8 PRELIMINARY; PRT; 121 AA.  
AC Q8YZW8;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein A110337.  
GN A110337.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21595285; PubMed-11759840;  
RA Kaneko T.; Nakamura Y.; Wolk C.P.; Kuritz T.; Sasamoto S.;  
RA Watanabe A.; Iitiguchi M.; Ishikawa A.; Kawashima K.; Kimura T.;  
RA Kishida Y.; Kohara M.; Matsumoto M.; Matsuno A.; Muraki A.;  
RA Nakazaki N.; Shimpo S.; Sugimoto M.; Takazawa M.; Yamada M.;  
RA Yasuda M.; Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003582; BAB72293.1; -;  
DR InterPro: IPR003477; Pfam: PF02452; Psmk; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 121 AA; 13266 MW; 8E01290F333B1B7A CRC64;  
Query Match  
Best Local Similarity 8.9%; Score 7; DB 16; Length 121;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 PVLSPFLR 31  
Db 56 PVLSPFLR 62  
RESULT 12  
Q8U038 PRELIMINARY; PRT; 146 AA.  
AC Q8U038;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE POL protein (Fragment).  
GN POL.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=98CMA307;  
RX MEDLINE-21565561; PubMed-11709100;  
RA Luk K.-C.; Kaplue L.; Zekeng L.; Soriano V.; Gurtler L.; Devare S.G.;  
RA Schochetman G.; Hackett J. Jr.;  
RT "Naturally Occurring Sequence Polymorphisms within HIV Type 1 Group O  
RT Protease.";  
RL AIDS Res. Hum. Retroviruses 17:1555-1561(2001).  
DR EMBL; AF380212; AAL32185.1; -;  
DR InterPro: IPR001995; Aspprotease\_ftrv.

```
DR InterPro: IPR001969; Aspprotease_site.
DR Pfam: PF00077; rvp: 1.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE: PS50175; ASP_PROT_RETROV; 1.
FT NON_TER 1
FT CHAIN 40 >138 PROTEASE.
FT NON_TER 146 146
SQ SEQUENCE 146 AA; 15018 MW; 70C5BA6539C7534 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 TSVPISP 57
   |||||
Db 7 TSVPISP 13

RESULT 13
O8BU36 PRELIMINARY; PRT; 146 AA.
AC O8BU36:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Pol protein (Fragment).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=98CMA376;
RX MEDLINE=21565561; PubMed=11709100;
RA Luk K.-C., Kapite L., Zekeng L., Soriano V., Gurtler L., Devare S.G.,
RA Schochetman G., Hackett J. Jr.;
RT "Naturally Occurring Sequence Polymorphisms within HIV Type 1 Group O
RT Protease.";
RL AIDS Res. Hum. Retroviruses 17:1555-1561(2001).
DR EMBL: AF380214; AAL32187.1; -.
DR InterPro: IPR001995; Aspprotease_rtrv.
DR PROSITE: IPR001969; Aspprotease_site.
DR Pfam: PF00077; rvp: 1.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE: PS50175; ASP_PROT_RETROV; 1.
FT NON_TER 1
FT CHAIN 40 >138 PROTEASE.
FT NON_TER 146 146
SQ SEQUENCE 146 AA; 15127 MW; F498137384AB00D6 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 TSVPISP 57
   |||||
Db 7 TSVPISP 13

RESULT 14
O9FMC7 PRELIMINARY; PRT; 178 AA.
AC O9FMC7:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 20.3 kDa protein.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned pl clones.";
RL DNA Res. 4:401-414(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carinci P., Chen H., Cheuk R., Hayaishizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene MUK1.15/AT5g04830 (GI:9758458).";
RT Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carinci P., Chen H., Cheuk R., Hayaishizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MUK1.15/AT5g04830 (GI:9758458).";
RT Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB008271; BAB08987.1; -.
DR EMBL: AF360335; AAK28632.1; -.
DR EMBL: AY051070; AAK93747.1; -.
KW Hypothetical protein.
SQ SEQUENCE 178 AA; 20294 MW; 955AC984A407FF54 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 SAFYSLP 44
   |||||
Db 66 SAFYSLP 72

RESULT 15
O9SPE1 PRELIMINARY; PRT; 200 AA.
AC O9SPE1:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 21.7 kDa protein (Fragment).
OS Amblyomma americanum.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Amblyomma.
OX NCBI_TaxID=6943;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-MIDGUT.
RA Jaworski D.C., Barbour A.G.;
RT Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF131719; AAK49812.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 200 200
SQ SEQUENCE 200 AA; 21688 MW; 6ABD8ADFCA3B84 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 4 SLPRPA 10  
| | | | |  
Db 124 SLPRPA 130

Search completed: March 4, 2003, 12:46:38  
Job time : 10.7249 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:43:52 ; Search time 6.34538 Seconds  
(without alignments)  
366.315 Million cell updates/sec

Title: US-09-234-208B-1

Sequence: 1 GTHSLPRPAAPVPLRMQP.....VGRGPDPAHVAVNLRYEG 79

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	79	4	US-09-630-155-1
2	79	100.0	419	4	US-09-630-155-2
3	7	8.9	328	4	US-09-300-672-2
4	7	8.9	635	2	US-09-014-969-11
5	7	8.9	1091	4	US-09-306-595C-7
6	6	7.6	12	2	US-08-811-492-146
7	6	7.6	15	1	US-08-036-555B-49
8	6	7.6	15	1	US-08-469-569-49
9	6	7.6	15	1	US-08-249-322A-49
10	6	7.6	15	1	US-08-469-526A-49
11	6	7.6	15	2	US-08-734-591A-49
12	6	7.6	15	2	US-08-469-660-49
13	6	7.6	15	3	US-08-341-018-82
14	6	7.6	15	4	US-08-470-335-49
15	6	7.6	15	4	US-08-735-021-49
16	6	7.6	15	4	US-08-734-664A-49
17	6	7.6	15	4	US-08-470-339-49
18	6	7.6	15	4	US-08-467-603-49
19	6	7.6	15	5	PCT-US94-05083C-46
20	6	7.6	15	5	PCT-US95-06846A-49
21	6	7.6	16	1	US-08-036-555B-37
22	6	7.6	16	1	US-08-469-569-37
23	6	7.6	16	1	US-08-249-322A-37
24	6	7.6	16	1	US-08-469-526A-37
25	6	7.6	16	2	US-08-734-591A-37
26	6	7.6	16	2	US-08-469-660-37
27	6	7.6	16	4	US-08-470-335-37

28	6	7.6	16	4	US-08-735-021-37	Sequence 37, Appl
29	6	7.6	16	4	US-08-734-664A-37	Sequence 37, Appl
30	6	7.6	16	4	US-08-470-339-37	Sequence 37, Appl
31	6	7.6	16	4	US-08-467-602-37	Sequence 37, Appl
32	6	7.6	16	5	PCT-US94-05083C-37	Sequence 37, Appl
33	6	7.6	16	5	PCT-US95-06846A-37	Sequence 37, Appl
34	6	7.6	34	1	US-08-118-270-81	Sequence 81, Appl
35	6	7.6	34	1	US-08-085-122-81	Sequence 11, Appl
36	6	7.6	34	2	US-08-319-052-21	Sequence 21, Appl
37	6	7.6	34	4	US-08-442-108B-21	Sequence 21, Appl
38	6	7.6	34	5	PCT-US93-08528-81	Sequence 81, Appl
39	6	7.6	74	4	US-09-134-001C-4092	Sequence 4092, Ap
40	6	7.6	145	1	US-08-565-386-15	Sequence 15, Appl
41	6	7.6	166	4	US-09-615-192A-273	Sequence 273, App
42	6	7.6	290	4	US-09-386-653A-7	Sequence 7, Appl
43	6	7.6	295	1	US-08-411-706-4	Sequence 4, Appl
44	6	7.6	304	4	US-09-088-651-2	Sequence 2, Appl
45	6	7.6	310	4	US-09-199-637A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-09-630-155-1  
; Sequence 1, Application US/09630155  
; Patent No. 6414130  
; GENERAL INFORMATION:  
; APPLICANT: Doherty, Joni Kristin and Gail M. Clinton  
; TITLE OF INVENTION: HER-2 BINDING ANTAGONISTS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVIS WRIGHT TREMAINE LLP  
; STREET: 1501 Fourth Avenue, 2600 Century Square  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: PC compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/630,155  
; FILING DATE: 16-Jan-2001  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Davison, Barry L.  
; REGISTRATION NUMBER: 47,309  
; REFERENCE/DOCKET NUMBER: 49321-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206 628-7621  
; TELEFAX: 206 628-7699  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 79  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: HER-2 ECD antagonist  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-630-155-1  
Query Match 100.0%; Score 79; DB 4; Length 79;  
Best Local Similarity 100.0%; Pred. No. 2.2e-69;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTHSLPRPAAPVPLRMQPGRAHPVLSFLRPSMDLVSAFYSLPLAPLSPTVPSVSV 60  
DB 1 GTHSLPRPAAPVPLRMQPGRAHPVLSFLRPSMDLVSAFYSLPLAPLSPTVPSVSV 60  
QY 61 GRGPDPAHVAVNLRYEG 79

Db 61 GRGPPDAHVAVNLSTRYEG 79

RESULT 2

US-09-630-155-2

Sequence 2, Application US/09630155

Patent No. 6414130

GENERAL INFORMATION:

APPLICANT: Doherty, Joni Kristin and Gail M. Clinton

TITLE OF INVENTION: HER-2 BINDING ANTAGONISTS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: DAVIS WRIGHT TREMAINE LLP

STREET: 1501 Fourth Avenue, 2600 Century Square

CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98101

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: PC compatible

OPERATING SYSTEM: Windows95

SOFTWARE: Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/630,155

FILING DATE: 16-Jan-2001

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Davison, Barry L.

REGISTRATION NUMBER: 47,309

REFERENCE/DOCKET NUMBER: 49321-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206 628-7621

TELEFAX: 206 628-7699

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 419

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: polypeptide

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-630-155-2

Query Match 100.0%; Score 79; DB 4; Length 419;  
Best Local Similarity 100.0%; Pred. No. 9.8e-69;

Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTHTLPRPAVVPPLRMQGPAPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPSV 60

Db 341 GTHTLPRPAVVPPLRMQGPAPVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPSV 400

QY 61 GRGPPDAHVAVNLSTRYEG 79

Db 401 GRGPPDAHVAVNLSTRYEG 419

RESULT 3

US-09-300-672-2

Sequence 2, Application US/09300672

Patent No. 6248937

GENERAL INFORMATION:

APPLICANT: Finkelstein, Ruth R.

APPLICANT: Lynch, Tim

APPLICANT: Goodman, Howard M.

APPLICANT: Wang, Ming-Li

TITLE OF INVENTION: A TRANSCRIPTION FACTOR REGULATING SEED DEVELOPMENT,

FILE REFERENCE: 480.89(HV)

CURRENT APPLICATION NUMBER: US/09/300,672

CURRENT FILING DATE: 1999-04-27

NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 2  
LENGTH: 328  
TYPE: PRT  
ORGANISM: Arabidopsis  
US-09-300-672-2

Query Match 8.9%; Score 7; DB 4; Length 328;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLPRPA 11

Db 144 LLPRPA 150

RESULT 4  
US-09-014-969-11

Sequence 11, Application US/09014969

Patent No. 5965397

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M.

APPLICANT: Lavallee, Edward R.

APPLICANT: Racie, Lisa A.

APPLICANT: Merberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Spaulding, Vikki

APPLICANT: Agostino, Michael J.

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/014,969

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 635 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-014-969-11

Query Match 8.9%; Score 7; DB 2; Length 635;  
Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 APLSPS 52

Db 9 APLSPS 15

RESULT 5



US-09-306-595C-7  
; Sequence 7, Application US/09306595C  
; Patent No. 6284506  
; GENERAL INFORMATION:  
; APPLICANT: HOSHINO, Tatsuo  
; APPLICANT: OIIMA, Kazuyuki  
; APPLICANT: SETOGUCHI, Yutaka  
; TITLE OF INVENTION: ISOPRENOID PRODUCTION  
; FILE REFERENCE: ISOPRENOID PRODUCTION  
; CURRENT APPLICATION NUMBER: US/09/306,595C  
; CURRENT FILING DATE: 1999-05-06  
; PRIOR APPLICATION NUMBER: 98108210  
; PRIOR FILING DATE: 1998-05-06  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 1091  
; TYPE: PRT  
; ORGANISM: Phaffia rhodozyma  
US-09-306-595C-7

Query Match  
Best local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 PISPSV 60  
DB 506 PISPSV 512

RESULT 6  
US-08-811-492-146  
; Sequence 146, Application US/08811492  
; Patent No. 5834247

; GENERAL INFORMATION:  
; APPLICANT: COMB, DONALD G.  
; APPLICANT: PERLER, FRANCINE B.  
; APPLICANT: JACK, WILLIAM E.  
; APPLICANT: XU, MING-OUN  
; APPLICANT: HODGES, ROBERT A.  
; APPLICANT: NOREN, CHRISTOPHER J.  
; APPLICANT: CHONG, SHAORONG S.C.  
; APPLICANT: ADAM, ERIC  
; APPLICANT: SOUTHWORTH, MAURICE  
; TITLE OF INVENTION: MODIFIED PROTEINS, METHODS OF THEIR  
; TITLE OF INVENTION: PRODUCTION AND METHODS FOR PURIFICATION OF TARGET  
; TITLE OF INVENTION: PROTEINS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GREGORY D. WILLIAMS; NEW ENGLAND BIOLABS, INC.  
; STREET: 32 TOZER ROAD  
; CITY: BEVERLY  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 01915

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC\DOS\MS\\_DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811,492  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/580,555  
FILING DATE: 29-DEC-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/496,247  
FILING DATE: 28-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/146,885  
FILING DATE: 03-NOV-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/004,139  
FILING DATE: 09-DEC-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams, Gregory D  
REGISTRATION NUMBER: 30901  
REFERENCE/DOCKET NUMBER: NEB-036C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 508-927-5054  
TELEFAX: 509-927-1705  
TELEX:  
INFORMATION FOR SEQ ID NO: 146:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-811-492-146

Query Match  
Best local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 LAPLSP 50  
DB 7 LAPLSP 12

RESULT 7  
US-08-036-555B-49  
; Sequence 49, Application US/08036555B  
; Patent No. 5530109

; GENERAL INFORMATION:  
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;  
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchionni, Mark;  
; APPLICANT: Chen, Maio Su; Hiles, Ian  
; TITLE OF INVENTION: Glial Mitogenic Factors, Their  
; TITLE OF INVENTION: Preparation and Use  
; NUMBER OF SEQUENCES: 184  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felle & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/036,555B  
FILING DATE: 24-MAR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,173  
FILING DATE: 23-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/940,389  
FILING DATE: 03-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,138  
FILING DATE: 30-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/863,703  
FILING DATE: 03-APRIL-1992  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: U.K. 91 07566.3  
FILING DATE: 10-APRIL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsai, Christine H.  
REGISTRATION NUMBER: 34,266  
REFERENCE/DOCKET NUMBER: LUD 5250.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-036-555B-49

Query Match 7.6%; Score 6; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SPVSVG 61  
|||||  
DB 2 SPVSVG 7

RESULT 8  
US-08-469-569-49  
Sequence 49, Application US/08469569  
Patent No. 5606032  
GENERAL INFORMATION:  
APPLICANT: Goodearl, Andrew; Stroobant, Paul;  
APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchionl, Mark;  
APPLICANT: Chen, Mao Su; Hiles, Ian  
TITLE OF INVENTION: Glial Mitogenic Factors, Their  
NUMBER OF INVENTION: Preparation and Use  
NUMBER OF SEQUENCES: 184  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felte & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,569  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/036,555  
FILING DATE: 24-MAR-1993  
APPLICATION NUMBER: 07/965,173  
FILING DATE: 23-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/940,389  
FILING DATE: 03-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,138  
FILING DATE: 30-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/863,703  
FILING DATE: 03-APRIL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.K. 91 07566.3  
FILING DATE: 10-APRIL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsai, Christine H.  
REGISTRATION NUMBER: 34,266

REFERENCE/DOCKET NUMBER: LUD 5250.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-469-569-49

Query Match 7.6%; Score 6; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SPVSVG 61  
|||||  
DB 2 SPVSVG 7

RESULT 9  
US-08-249-322A-49  
Sequence 49, Application US/08249322A  
Patent No. 5716930  
GENERAL INFORMATION:  
APPLICANT: Goodearl, Andrew; Stroobant, Paul;  
APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchionl, Mark;  
APPLICANT: Chen, Mao Su; Hiles, Ian  
TITLE OF INVENTION: Glial Mitogenic Factors, Their  
NUMBER OF INVENTION: Preparation and Use  
NUMBER OF SEQUENCES: 184  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felte & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/249,322A  
FILING DATE: 26-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/036,555  
FILING DATE: 24-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,173  
FILING DATE: 23-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/940,389  
FILING DATE: 03-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,138  
FILING DATE: 30-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/863,703  
FILING DATE: 03-APRIL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.K. 91 07566.3  
FILING DATE: 10-APRIL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsai, Christine H.  
REGISTRATION NUMBER: 34,266  
REFERENCE/DOCKET NUMBER: LUD 250.4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884

;; INFORMATION FOR SEQ ID NO: 49:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 15  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
US-08-249-322A-49

Query Match 7.6%; Score 6; DB 1;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SPVSVG 61  
|||||  
Db 2 SPVSVG 7

RESULT 10  
US-08-469-526A-49  
; Sequence 49, Application US/08469526A  
; Patent No. 5792849  
; GENERAL INFORMATION:  
; APPLICANT: Goodearl, Andrew  
; APPLICANT: Stroobant, Paul  
; APPLICANT: Minghetti, Luisa  
; APPLICANT: Waterfield, Michael  
; APPLICANT: Marchionni, Mark  
; APPLICANT: Chen, Maio Su  
; APPLICANT: Hiles, Ian  
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
; TITLE OF INVENTION: PREPARATION AND USE  
; NUMBER OF SEQUENCES: 187  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Clark & Elbing LLP  
; STREET: 176 Federal Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,526A  
; FILING DATE: 06 June 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/036,555  
; FILING DATE: 24-MAR-1993  
; APPLICATION NUMBER: 07/965,173  
; FILING DATE: 23-OCT-1992  
; APPLICATION NUMBER: 07/940,389  
; FILING DATE: 03-SEP-1992  
; APPLICATION NUMBER: 07/907,138  
; FILING DATE: 03-JUN-1992  
; APPLICATION NUMBER: 07/863,703  
; FILING DATE: 03-APRIL-1992  
; APPLICATION NUMBER: U.K. 91 07566.3  
; FILING DATE: 10-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bieker-Brady, Kristina  
; REGISTRATION NUMBER: 39,109  
; REFERENCE/DOCKET NUMBER: 04585/00200A  
; TELEPHONE: 617-428-0200  
; TELEFAX: 617-428-7045  
; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15  
; TYPE: amino acid  
; STRANDEDNESS: single

;; TOPOLOGY: linear  
US-08-469-526A-49

Query Match 7.6%; Score 6; DB 1;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SPVSVG 61  
|||||  
Db 2 SPVSVG 7

RESULT 11  
US-08-734-591A-49  
; Sequence 49, Application US/08734591A  
; Patent No. 5854220  
; GENERAL INFORMATION:  
; APPLICANT: Goodearl, Andrew  
; APPLICANT: Stroobant, Paul  
; APPLICANT: Minghetti, Luisa  
; APPLICANT: Waterfield, Michael  
; APPLICANT: Marchionni, Mark  
; APPLICANT: Hiles, Ian  
; APPLICANT: Chen, Mario  
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
; TITLE OF INVENTION: PREPARATION AND USE  
; NUMBER OF SEQUENCES: 187  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Clark & Elbing LLP  
; STREET: 176 Federal Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM Compatible Pentium  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: Wordperfect (Version 7.0)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/734,591A  
; FILING DATE: 22-OCT-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/470,335  
; FILING DATE: 06-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/036,555  
; FILING DATE: 03-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/965,173  
; FILING DATE: 23-OCT-1992  
; APPLICATION NUMBER: 07/940,389  
; FILING DATE: 03-SEP-1992  
; APPLICATION NUMBER: 07/907,138  
; FILING DATE: 30-JUN-1992  
; APPLICATION NUMBER: 07/863,703  
; FILING DATE: 03-APR-1992  
; APPLICATION NUMBER: UK 91 07566.3  
; FILING DATE: 10-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bieker-Brady, Kristina  
; REGISTRATION NUMBER: 39,109  
; REFERENCE/DOCKET NUMBER: 04585/00200P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 428-0200  
; TELEFAX: (617) 428-7045  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 49:

SEQUENCE CHARACTERISTICS:  
LENGTH: 15  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-734-591A-49

Query Match 7.6%; Score 6; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SPVSVG 61  
|||||  
Db 2 SPVSVG 7

RESULT 12  
US-08-469-660-49  
; Sequence 49, Application US/08469660  
; Patent No. 5876973

GENERAL INFORMATION:  
APPLICANT: Gwynne, David I.; Marchionni, Mark;  
APPLICANT: McBurney, Robert N.  
TITLE OF INVENTION: INHIBITORS OF CELL PROLIFERATION,  
TITLE OF INVENTION: THEIR PREPARATION AND USE  
NUMBER OF SEQUENCES: 184  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
ZIP: 0211-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,660  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/011,396  
FILING DATE: 29-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/984,085  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/951,747  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/927,337  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 04585/017004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: 200154  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-469-660-49

Query Match 7.6%; Score 6; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SPVSVG 61

Db 2 SPVSVG 7  
|||||

RESULT 13  
US-08-341-018-82  
; Sequence 82, Application US/08341018A  
; Patent No. 6087323

GENERAL INFORMATION:  
APPLICANT: Gwynne, David I.  
APPLICANT: Mahanthappa, Nagesh K.  
APPLICANT: Marchionni, Mark A.  
APPLICANT: Birmingham-McDonogh, Olivia  
APPLICANT: Goldin, Stanley M.  
APPLICANT: McBurney, Robert N.

TITLE OF INVENTION: USE OF NEUREGULINS AS MODULATORS OF  
TITLE OF INVENTION: CELLULAR COMMUNICATION  
FILE REFERENCE: 04585/041001  
CURRENT APPLICATION NUMBER: US/08/341,018A  
CURRENT FILING DATE: 1994-11-17  
NUMBER OF SEQ ID NOS: 87  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 82  
LENGTH: 15  
TYPE: PRP  
ORGANISM: Bos taurus  
US-08-341-018-82

Query Match 7.6%; Score 6; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SPVSVG 61  
|||||  
Db 2 SPVSVG 7

RESULT 14  
US-08-470-335-49

; Sequence 49, Application US/08470335F  
; Patent No. 6147190  
GENERAL INFORMATION:  
APPLICANT: GOODEARL, ANDREW  
APPLICANT: STROOBANT, PAUL  
APPLICANT: MINGHETTI, LUISA  
APPLICANT: WATERFIELD, MICHAEL  
APPLICANT: MARCHIONNI, MARK  
APPLICANT: CHEN, MARIO S.  
APPLICANT: HILES, IAN  
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
TITLE OF INVENTION: PREPARATION AND USE  
FILE REFERENCE: 04585/00200B  
CURRENT APPLICATION NUMBER: US/08/470,335F  
CURRENT FILING DATE: 1995-06-06  
EARLIER APPLICATION NUMBER: 08/036,555  
EARLIER FILING DATE: 1993-03-24  
NUMBER OF SEQ ID NOS: 252  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 49  
LENGTH: 15  
TYPE: PRP  
ORGANISM: Bos taurus  
US-08-470-335-49

Query Match 7.6%; Score 6; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SPVSVG 61  
|||||  
Db 2 SPVSVG 7

```
RESULT 15
US-08-735-021-49
; Sequence 49, Application US/08735021B
; Patent No. 6194377
; GENERAL INFORMATION:
; APPLICANT: GOODEARL, ANDREW
; APPLICANT: STROBANT, PAUL
; APPLICANT: MINGHETTI, LUISA
; APPLICANT: WATERFIELD, MICHAEL
; APPLICANT: MARCHIONNI, MARK
; APPLICANT: CHEN, MARIO S.
; APPLICANT: HILES, IAN
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; FILE REFERENCE: 04585/00200L
; CURRENT APPLICATION NUMBER: US/08/735,021B
; EARLIER APPLICATION NUMBER: 08/472,065
; EARLIER FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/036,555
; EARLIER FILING DATE: 1993-03-24
; EARLIER APPLICATION NUMBER: 07/965,173
; EARLIER FILING DATE: 1992-10-23
; EARLIER APPLICATION NUMBER: 07/940,389
; EARLIER FILING DATE: 1992-09-03
; EARLIER APPLICATION NUMBER: 07/907,138
; EARLIER FILING DATE: 1992-06-30
; EARLIER APPLICATION NUMBER: 07/863,703
; EARLIER FILING DATE: 1992-04-03
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 49
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Bos taurus
US-08-735-021-49

Query Match          7.6%; Score 6; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 SPVSVG 61
   |||||
Db 2 SPVSVG 7
```

Search completed: March 4, 2003, 12:48:27  
Job time : 7.34538 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:46:47 ; Search time 3.3313 Seconds  
(without alignments)  
1000.035 Million cell updates/sec

Title: US-09-234-208B-1  
Perfect score: 79  
Sequence: 1 GTHSLPRPAAVPVPLRMQP.....VGRCPDAHVAVNLSRYEG 79

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 188354 seqs, 42170167 residues

Word size : 0

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCT05\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	8.9	615	10	US-09-864-761-44740
2	7	8.9	215	9	US-10-001-876-210
3	7	8.9	824	9	US-09-909-567B-53
4	7	8.9	880	9	US-09-893-519A-36
5	7	8.9	906	10	US-09-746-491-48
6	7	7.6	30	10	US-09-864-761-45510
7	6	7.6	34	9	US-09-755-109-21
8	6	7.6	38	10	US-09-864-761-41953
9	6	7.6	68	10	US-09-764-887-226
10	6	7.6	77	10	US-09-864-761-35919
11	6	7.6	86	10	US-09-764-869-1118
12	6	7.6	93	10	US-09-764-869-798
13	6	7.6	114	10	US-09-864-761-38224
14	6	7.6	131	10	US-09-771-161A-148
15	6	7.6	147	10	US-09-764-847-828
16	6	7.6	168	9	US-09-738-626-5722
17	6	7.6	168	10	US-09-925-301-1470
18	6	7.6	238	12	US-10-024-579-16
19	6	7.6	243	9	US-09-989-442-122

20	6	7.6	257	12	US-10-024-579-14	Sequence 14, Appl
21	6	7.6	264	10	US-09-848-696-4	Sequence 4, Appl
22	6	7.6	264	12	US-10-024-579-12	Sequence 12, Appl
23	6	7.6	267	9	US-09-808-602-23	Sequence 23, Appl
24	6	7.6	267	9	US-09-808-602-25	Sequence 25, Appl
25	6	7.6	283	12	US-10-024-579-10	Sequence 10, Appl
26	6	7.6	290	9	US-10-041-006A-7	Sequence 7, Appl1
27	6	7.6	290	9	US-10-028-072-222	Sequence 222, App
28	6	7.6	290	9	US-10-121-049-222	Sequence 222, App
29	6	7.6	290	9	US-10-123-904-222	Sequence 222, App
30	6	7.6	290	9	US-10-140-470-222	Sequence 222, App
31	6	7.6	290	9	US-10-175-746-222	Sequence 222, App
32	6	7.6	290	9	US-10-176-918-222	Sequence 222, App
33	6	7.6	290	9	US-10-176-921-222	Sequence 222, App
34	6	7.6	290	9	US-10-137-865-222	Sequence 222, App
35	6	7.6	290	9	US-10-140-474-222	Sequence 222, App
36	6	7.6	290	9	US-10-142-431-222	Sequence 222, App
37	6	7.6	290	9	US-10-143-114-222	Sequence 222, App
38	6	7.6	290	9	US-10-140-002-222	Sequence 222, App
39	6	7.6	290	12	US-10-040-655-7	Sequence 7, Appl
40	6	7.6	298	10	US-09-764-853-552	Sequence 552, App
41	6	7.6	299	9	US-09-808-602-63	Sequence 63, Appl
42	6	7.6	306	12	US-10-080-644-11	Sequence 11, Appl
43	6	7.6	310	9	US-09-975-719-11	Sequence 49, Appl
44	6	7.6	316	10	US-09-799-777-49	Sequence 160, App
45	6	7.6	323	9	US-09-984-245-160	

#### ALIGNMENTS

RESULT 1  
US-09-864-761-44740  
Sequence 44740, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aomicla-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 44740
; LENGTH: 61
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC016057.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.72
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.59
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.52
; OTHER INFORMATION: SWISSPROT HIT: Q13563, EVALUATE 2.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BF570694.1, EVALUATE 4.00e-16
US-09-864-761-44740
```

```

Query Match      8.9%; Score 7; DB 10; Length 61;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 23 AHPVLSF 29
      |||||
DB 37 AHPVLSF 43
```

```

RESULT 2
US-10-001-876-210
; Sequence 210, Application US/10001876
; Patent No. US20020177140A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Caferkey, Robert
; APPLICANT: Ali, Shujath
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenchua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and
; FILE REFERENCE: DEX-0285
; CURRENT APPLICATION NUMBER: US/10/001,876
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,186
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 210
; LENGTH: 215
; TYPE: PRF
; ORGANISM: Homo sapien
US-10-001-876-210
```

```

Query Match      8.9%; Score 7; DB 9; Length 215;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 27 LSFRLPS 33
      |||||
DB 60 LSFRLPS 66
```

```

RESULT 3
US-09-909-567B-53
; Sequence 53, Application US/09909567B
; Publication No. US2003002257A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Macina, Roberto A.
; APPLICANT: Nair, Manoj
; APPLICANT: Chen, Seiyu
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes
; FILE REFERENCE: DEX-0214
; CURRENT APPLICATION NUMBER: US/09/909,567B
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 60/219,834
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 824
; TYPE: PRF
; ORGANISM: Homo sapien
US-09-909-567B-53
```

```

Query Match      8.9%; Score 7; DB 9; Length 824;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 11 AHPVPLR 17
      |||||
DB 405 AHPVPLR 411
```

```

RESULT 4
US-09-893-519A-36
; Sequence 36, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BOURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESTILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAO, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTI-FUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/16548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 880
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 109
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/AAC39727
; DATABASE ENTRY DATE: 1998-05-06
; RELEVANT RESIDUES: (1)..(880)
US-09-893-519A-36
```

```

Query Match      8.9%; Score 7; DB 9; Length 880;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



QY 24 HPVLSFL 30  
|||||||  
DB 424 HPVLSFL 430

RESULT 5  
US-09-746-491-48  
; Sequence 48, Application US/09746491  
; Patent No. US20020137202A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Catherine E.  
; TITLE OF INVENTION: No. US20020137202A1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-621  
; CURRENT APPLICATION NUMBER: US/09/746,491  
; CURRENT FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: USSN 60/171,329  
; PRIOR FILING DATE: 1999-12-21  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 48  
; LENGTH: 906  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-746-491-48

Query Match 8.9%; Score 7; DB 10; Length 906;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 VAVNLSR 76  
|||||||  
DB 117 VAVNLSR 123

RESULT 6  
US-09-864-761-45510  
; Sequence 45510, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 45510  
; LENGTH: 30  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC010721.2  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.66  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.61  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.63  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.59  
; OTHER INFORMATION: EST\_HUMAN HIT: BF672312.1, EVALUE 8.00e-04  
US-09-864-761-45510

Query Match 7.6%; Score 6; DB 10; Length 30;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 RGPDPD 67  
|||||||  
DB 12 RGPDPD 17

RESULT 7  
US-09-755-109-21  
; Sequence 21, Application US/09755109  
; Publication No. US20030027981A1  
; GENERAL INFORMATION:  
; APPLICANT: DANO, KEID  
; BLASI, FRANCESCO  
; ROLDAN, ANN LOURING  
; CUBELLIS, MARIA VITTORIA  
; MASUCCI, MARIA TERESA  
; APPELLA, ETTORE  
; SCHLEUNING, WOLF-DIETER  
; BEHRENDT, NIELS  
; RONNE, EBBE  
; KRISTENSEN, PETER  
; TITLE OF INVENTION: UROKINASE-TYPE PLASMINOGEN ACTIVATOR  
; RECEPTOR  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 SEVENTH STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/755,109  
; FILING DATE: 08-Jan-2001  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/635,666

FILING DATE: 22-APR-1996  
APPLICATION NUMBER: US 08/319,052  
FILING DATE: 06-OCT-1994  
APPLICATION NUMBER: US 07/824,189  
FILING DATE: 06-DEC-1991  
APPLICATION NUMBER: WO PCT/JP90/00090  
FILING DATE: 18-OCT-1990  
APPLICATION NUMBER: US 07/374,854  
FILING DATE: 03-JUL-1989  
APPLICATION NUMBER: US 07/334,613  
FILING DATE: 07-APR-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: DANO-1F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 21:  
US-09-755-109-21

Query Match 7.6%; Score 6; DB 9; Length 34;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 OGPAP 24  
|||||  
DB 8 OGPAP 13

RESULT 8  
US-09-864-761-41953  
Sequence 41953, Application US/09864761  
Patent No. US2002048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 41953  
LENGTH: 38  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC023344.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.8  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8  
US-09-864-761-41953

Query Match 7.6%; Score 6; DB 10; Length 38;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 PLSPTS 52  
|||||  
DB 24 PLSPTS 29

RESULT 9  
US-09-764-887-226  
Sequence 226, Application US/09764887  
Patent No. US20020042096A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PA113  
CURRENT APPLICATION NUMBER: US/09/764,887  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 658  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 226  
LENGTH: 68  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (6)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-887-226

Query Match 7.6%; Score 6; DB 10; Length 68;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 55 ISPSV 60  
|||||  
DB 28 ISPSV 33

```

RESULT 10
US-09-864-761-35919
; Sequence 35919, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35919
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008013.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.7
; OTHER INFORMATION: EST_HUMAN HIT: R09295.1, EVALUATE 1.10e+00
US-09-864-761-35919
Query Match 7.6%; Score 6; DB 10; Length 77;
Best Local Similarity 100.0%; Pred. No. 66;

```

```

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 LPLRPA 10
|11111
Db 9 LPLRPA 14

RESULT 11
US-09-764-869-1118
; Sequence 1118, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - REFER TO PALM OR FILE WRAPPER
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1118
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (48)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-1118
Query Match 7.6%; Score 6; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 LPLAPL 48
|11111
Db 61 LPLAPL 66

RESULT 12
US-09-764-869-798
; Sequence 798, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - REFER TO PALM OR FILE WRAPPER
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 798
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (56)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (60)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (79)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (83)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-798

```

Query Match 7.6%; Score 6; DB 10; Length 93;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 SVGRGP 64  
|||||  
DB 37 SVGRGP 42

## RESULT 13

US-09-864-761-38224  
; Sequence 38224, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmics-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 38224  
; LENGTH: 114  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL034384.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.86  
; OTHER INFORMATION: SWISSPROT HIT: P08503, EVALUATE 2.70e-01  
; OTHER INFORMATION: EST\_HUMAN HIT: AV703581.1, EVALUATE 2.10e+00  
US-09-864-761-38224

Query Match 7.6%; Score 6; DB 10; Length 114;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 51 TSVPIS 56  
|||||  
DB 73 TSVPIS 78

## RESULT 14

US-09-771-161A-148  
; Sequence 148, Application US/09771161A  
; Patent No. US20020110811A1  
; GENERAL INFORMATION:  
; APPLICANT: LEVINE, et al.  
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
; FILE REFERENCE: 802620-2005.1  
; CURRENT APPLICATION NUMBER: US/09/771,161A  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 09/724,676  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 136776  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 135619  
; PRIOR FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 273  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 148  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-771-161A-148

Query Match 7.6%; Score 6; DB 10; Length 131;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 THSLIP 7  
|||||  
DB 40 THSLIP 45

## RESULT 15

US-09-764-847-828  
; Sequence 828, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764,847  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 828  
; LENGTH: 147  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-847-828

Query Match 7.6%; Score 6; DB 10; Length 147;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LLPRPA 10  
|||||  
DB 108 LLPRPA 113

Search completed: March 4, 2003, 12:54:25  
Job time : 4.33133 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:36:41 ; Search time 53.8474 Seconds  
(without alignments)  
1036.856 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 419

Sequence: 1 METALCRWGLLALLPPGA.....VGRGPPDAHVAVNLSRYEG 419

Scoring table: OLIGO

Gapop 60.0 , Gapept 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A.Geneseq\_101002:\*

```
1: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1980.DAT:*
2: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1981.DAT:*
3: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1982.DAT:*
4: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1983.DAT:*
5: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1984.DAT:*
6: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1985.DAT:*
7: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1986.DAT:*
8: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1987.DAT:*
9: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1988.DAT:*
10: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1989.DAT:*
11: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1990.DAT:*
12: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1991.DAT:*
13: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1992.DAT:*
14: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1993.DAT:*
15: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1994.DAT:*
16: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1995.DAT:*
17: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1996.DAT:*
18: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1997.DAT:*
19: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1998.DAT:*
20: /SIDs2/gcgdata/geneseq/genesep-emb1/AA1999.DAT:*
21: /SIDs2/gcgdata/geneseq/genesep-emb1/AA2000.DAT:*
22: /SIDs2/gcgdata/geneseq/genesep-emb1/AA2001.DAT:*
23: /SIDs2/gcgdata/geneseq/genesep-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	340	81.1	645	22	Human ErbB2 oncopr
2	340	81.1	645	22	Human ErbB2 extrac
3	340	81.1	653	21	Human ErbB2 extrac
4	340	81.1	653	23	Human ErbB2 extrac
5	340	81.1	712	21	Human ErbB2/neu on
6	340	81.1	712	23	Human ErbB2/neu on
7	340	81.1	782	18	Human ErbB2/neu on
8	340	81.1	919	21	Human ErbB2/neu on
9	340	81.1	919	23	Human ErbB2/neu on
10	340	81.1	1200	21	Human ErbB2/neu on

11	340	81.1	1255	17	AAW01111
12	340	81.1	1255	20	AAW92406
13	340	81.1	1255	21	AAW21198
14	340	81.1	1255	21	AAW84780
15	340	81.1	1255	21	AAW92620
16	340	81.1	1255	22	AAE12130
17	340	81.1	1255	22	AAW85458
18	340	81.1	1255	22	AAW88267
19	340	81.1	1255	22	AAW60167
20	340	81.1	1255	23	AAE24067
21	340	81.1	1255	23	AAE20479
22	340	81.1	1255	23	AAW51143
23	340	81.1	1255	23	AAU77114
24	340	81.1	1255	23	AAU74545
25	319	76.1	951	21	AAW44993
26	307	73.3	1223	23	AAU98923
27	292	69.7	479	22	AAE13112
28	292	69.7	555	22	AAE13108
29	292	69.7	564	22	AAE13110
30	292	69.7	690	22	AAE13109
31	292	69.7	697	22	AAE13111
32	289	69.0	289	22	AAE13120
33	191	45.6	191	19	AAW53354
34	191	45.6	191	20	AAW14572
35	191	45.6	191	22	AAW48767
36	187	44.6	1433	14	AAW39568
37	166	39.6	166	19	AAW59345
38	166	39.6	166	22	AAW62074
39	166	39.6	166	22	AAW48763
40	96	22.9	97	17	AAW00327
41	86	20.5	624	11	AAW08222
42	83	19.8	419	22	AAE09181
43	83	19.8	419	22	AAE09183
44	83	19.8	419	22	AAE09200
45	83	19.8	419	22	AAE09202

#### ALIGNMENTS

RESULT 1	AAW60408	standard; Protein; 645 AA.
ID	AAW60408	
AC	AAW60408	
XX		
DT	24-APR-2001	(first entry)
XX		
DE	Human ErbB2 oncoprotein, SEQ ID NO:13.	
XX		
KW	Anti-ErbB2 monoclonal antibody 2C4; HER2; mouse; murine; humanised; VL;	
KW	light chain variable region; cancer; cytostatic; EGFR-expressing cancer;	
KW	epidermal growth factor receptor; colon cancer; rectal cancer; tumour;	
KW	colorectal cancer; non-small cell lung cancer; metastatic breast cancer;	
XX	affinity purification.	
OS	Homo sapiens.	
XX		
PN	WO200100245-A2.	
XX		
PD	04-JAN-2001.	
XX		
PF	23-JUN-2000; 2000MO-US17366.	
XX		
PR	25-JUN-1999; 99US-0141316.	
XX		
PA	(GETH ) GENENTECH INC.	
XX		
PI	Adams CW, Presta LG, Sliwkowsky M;	
XX		
DR	WPI; 2001-080862/09.	
XX		
PT	Treating cancer in a human, where the cancer expresses epidermal growth	

PT factor receptor (EGFR), comprises administering an antibody which binds  
PT ErbB2 -

PS Example 1; Fig 1A: 89pp; English.

CC The invention relates to a method for treating cancer in a human patient,  
CC wherein the cancer expresses epidermal growth factor receptor (EGFR),  
CC comprising administering an antibody which binds ErbB2 (HER2; AAB60408).  
CC In particular, the anti-ErbB2 antibody is the murine monoclonal antibody  
CC 2C4 (AAB60396, AAB60397) or a humanised version of 2C4 (AAB60398,  
CC AAB60399). The invention also encompasses an isolated nucleic acid  
CC encoding a humanised ErbB2-binding antibody; vectors and host cells  
CC comprising such nucleic acids; the recombinant production of a humanised  
CC ErbB2-binding antibody; and an immunocjugate comprising a humanised  
CC ErbB2-binding antibody and a cytotoxic drug. The ErbB2-binding antibodies  
CC act by antagonising ErbB receptors, and as inhibitors of transforming  
CC growth factor alpha (TGF-alpha)-activated mitogen activated protein  
CC kinase (MAPK). The method of the invention is used for treating cancer,  
CC especially colon cancer, rectal cancer, colorectal cancer, lung cancer  
CC (especially non-small cell lung cancer), or breast cancer (especially  
CC metastatic breast cancer). The antibodies may also have non-therapeutic  
CC uses e.g., as affinity purification agents. Using an antibody which binds  
CC to ErbB2 to treat cancer is preferable to the use of EGFR-targeted  
CC drugs, as EGFR is also highly expressed in other tissues such as the  
CC liver and skin, where the active drug will also bind, with skin toxicity  
CC having been observed for EGFR-targeted drugs. Antibodies which bind  
CC ErbB2 are anticipated to have a better safety profile than such drugs.  
CC The present sequence represents human ErbB2.

XX Sequence 645 AA:

Query Match 81.1%; Score 340; DB 22; Length 645;  
Best Local Similarity 100.0%; Pred. No. 3.9e-302;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METALCRNGLLIALPPGAASVOVCTGDMKLRPASPEHMDIRHLHYGCGVVGML 60  
DB 1 METALCRNGLLIALPPGAASVOVCTGDMKLRPASPEHMDIRHLHYGCGVVGML 60  
QY 61 ELTYLPTNLSFLDIOEVGVYLAHNOVROPLORLIRIVRGTOLEFEDNALAVLDNG 120  
DB 61 ELTYLPTNLSFLDIOEVGVYLAHNOVROPLORLIRIVRGTOLEFEDNALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTIIMKDIFFHKNNOIA 180  
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTIIMKDIFFHKNNOIA 180  
QY 181 LTLIDNRSRACHPCSPMKSGRCWGESSESDCSLTRVCAGGACARCKGPLPTDCCHEOC 240  
DB 181 LTLIDNRSRACHPCSPMKSGRCWGESSESDCSLTRVCAGGACARCKGPLPTDCCHEOC 240  
QY 241 AAGCTPRKSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRTTFGASCTYACP 300  
DB 241 AAGCTPRKSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRTTFGASCTYACP 300  
QY 301 YNVLSTDVGSCTLVCPDLHNOEVTAEEDGTORCEKSKPCAR 340  
DB 301 YNVLSTDVGSCTLVCPDLHNOEVTAEEDGTORCEKSKPCAR 340

RESULT 2  
AAB61593  
ID AAB61593 standard; protein; 645 AA.

XX AAB61593;

XX 04-APR-2001 (first entry)

XX Human ErbB2 extracellular domain.

XX Human; ErbB2; cytosolic; prostate cancer; receptor tyrosine kinase;  
KW antibody; ErbB receptor; monoclonal antibody 2C4; variable light chain.  
XX

OS Homo sapiens.

XX WO200100238-A1.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-US17423.

XX 25-JUN-1999; 99US-0141315.

XX (GENTH ) GENENTECH INC.

XX (SLOK ) SLOAN KETTERING INST CANCER RES.

XX Agus DB, Scher HI, Sliwkowski MX;

XX WPI: 2001-159131/16.

XX Treating prostate cancer in a human comprises administering an antibody  
XX which binds ErbB2 and blocks ligand activation of an ErbB receptor -

PS Disclosure: Fig 1; 93pp; English.

CC The ErbB family of receptor tyrosine kinases are important mediators of  
CC cell growth, differentiation and survival. The receptor family includes  
CC four distinct members including Epidermal Growth Factor Receptor (EGFR or  
CC ErbB1), HER2 (ErbB2 or p185<sup>neu</sup>), HER3 (ErbB3) and Her4 (ErbB4 or tyro2).  
CC The present invention relates to a method for treating prostate cancer.  
CC The method comprises administering an antibody which binds ErbB2 and  
CC blocks ligand activation of an ErbB receptor. Preferably, the antibody  
CC blocks binding of monoclonal antibody 2C4 to ErbB2 and/or blocks  
CC TGF-alpha activation of mitogen-activated protein kinase (MAPK). The  
CC present sequence is the extracellular domain of human ErbB2.

XX Sequence 645 AA:

Query Match 81.1%; Score 340; DB 22; Length 645;  
Best Local Similarity 100.0%; Pred. No. 3.9e-302;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METALCRNGLLIALPPGAASVOVCTGDMKLRPASPEHMDIRHLHYGCGVVGML 60  
DB 1 METALCRNGLLIALPPGAASVOVCTGDMKLRPASPEHMDIRHLHYGCGVVGML 60  
QY 61 ELTYLPTNLSFLDIOEVGVYLAHNOVROPLORLIRIVRGTOLEFEDNALAVLDNG 120  
DB 61 ELTYLPTNLSFLDIOEVGVYLAHNOVROPLORLIRIVRGTOLEFEDNALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTIIMKDIFFHKNNOIA 180  
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTIIMKDIFFHKNNOIA 180  
QY 181 LTLIDNRSRACHPCSPMKSGRCWGESSESDCSLTRVCAGGACARCKGPLPTDCCHEOC 240  
DB 181 LTLIDNRSRACHPCSPMKSGRCWGESSESDCSLTRVCAGGACARCKGPLPTDCCHEOC 240  
QY 241 AAGCTPRKSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRTTFGASCTYACP 300  
DB 241 AAGCTPRKSDCLACHFNHSGICELHCPALVYNTDFESMPNPEGRTTFGASCTYACP 300  
QY 301 YNVLSTDVGSCTLVCPDLHNOEVTAEEDGTORCEKSKPCAR 340  
DB 301 YNVLSTDVGSCTLVCPDLHNOEVTAEEDGTORCEKSKPCAR 340

RESULT 3  
AAB21200  
ID AAB21200 standard; protein; 653 AA.

XX AAB21200;

XX 12-JAN-2001 (first entry)

XX Extracellular HER-2/neu protein.



```

XX  HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
KM  breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW  colon cancer.
XX
XX  Unidentified.
OS
PN  WO200044899-A1.
PD  03-AUG-2000.
XX
PF  28-JAN-2000; 2000WO-US02164.
XX
PR  29-JAN-1999; 99GS-0117976.
XX
PA  (CORI-) CORIXA CORP.
PA  (SMK ) SMITHKLINE BEECHAM.
XX
PI  Cheever MA, Gheysen D;
DR  WPI; 2000-505976/45.
XX
PT  HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT  useful for vaccinating against breast, ovarian, colon, lung and
PT  prostate cancers -
XX
PS  Claim 2; Fig 9; 128pp; English.
XX
CC  The present sequence is the extracellular HER-2/neu protein. HER-2/neu is
CC  a member of the tyrosine kinase family of receptor-like glycoproteins and
CC  shows homology to the epidermal growth factor receptor (EGFR). It
CC  probably plays a part in cell growth and/or differentiation. The
CC  HER-2/neu gene is an oncogene. An HER-2/neu fusion protein comprising a
CC  HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC  domain may be used to treat or prevent cancer by eliciting or
CC  enhancing an immune response to the HER-2/neu protein. It may be used
CC  to treat malignancies such as breast, ovarian, colon, lung and
CC  prostate cancers, and may be used as an antigen to vaccinate against
CC  these neoplasias.
XX
SQ  Sequence 653 AA:

Query Match      81.1%; Score 340; DB 21; Length 653;
Best Local Similarity 100.0%; Pred. No. 4e-302;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 MELAALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHYOGCQVVOGNTL 60
DB  1 MELAALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHYOGCQVVOGNTL 60
OY  61 ELTYLPTNASTSLFLDIOEVQGVYLAHNOVROVPLQRLRIYRGTOLEFDNTVALAVLDNG 120
DB  61 ELTYLPTNASTSLFLDIOEVQGVYLAHNOVROVPLQRLRIYRGTOLEFDNTVALAVLDNG 120
OY  121 DELNNTPTVTSASPGSLRELOLRSLTEILKGVLTDRNQLCYQDTIILKKDIFHKNNQIA 180
DB  121 DELNNTPTVTSASPGSLRELOLRSLTEILKGVLTDRNQLCYQDTIILKKDIFHKNNQIA 180
OY  181 LFLIDTNRSRACHPCSPMKSGSRCSWESSESDCSLRTYCAGGACARCKPLPDDCCHQC 240
DB  181 LFLIDTNRSRACHPCSPMKSGSRCSWESSESDCSLRTYCAGGACARCKPLPDDCCHQC 240
OY  241 AAGCTGPKHSDDLACLFHNSGICELHCPALVYNTDFESMPNPEGRYTFGASCVTACP 300
DB  241 AAGCTGPKHSDDLACLFHNSGICELHCPALVYNTDFESMPNPEGRYTFGASCVTACP 300
OY  301 YNVLSTDVGSCTLVCPHLHNOEYTAEDGTQRCERCKSPCAR 340
DB  301 YNVLSTDVGSCTLVCPHLHNOEYTAEDGTQRCERCKSPCAR 340

```

```

ID  AAM51145 standard; Protein; 653 AA.
XX
AC  AAM51145;
XX
DT  17-JUN-2002 (first entry)
XX
DE  Human Her-2/neu oncoprotein extracellular domain.
XX
KM  Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW  tyrosine kinase; receptor; c-erbB2; gene therapy.
XX
OS  Homo sapiens.
PN  WO200212341-A2.
XX
PD  14-FEB-2002.
XX
PF  03-AUG-2001; 2001WO-US24283.
XX
PR  03-AUG-2000; 2000US-0632507.
XX
PA  (CORI-) CORIXA CORP.
PA  (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI  Cheever MA, Gheysen D;
DR  WPI; 2002-241743/29.
XX
PT  Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT  or enhancing an immune response to the protein, has Her-2/neu
PT  extracellular domain fused to Her-2/neu intracellular or
PT  phosphorylation domain -
XX
PS  Claim 2; Fig 9; 141pp; English.
XX
CC  The present sequence is that of the extracellular domain of
CC  human Her-2/neu (p185 glycoprotein or c-erbB2), an oncogenic
CC  self-protein and target for anti-cancer vaccines. The Her-2/neu
CC  gene is amplified and p185 is overexpressed in a variety of cancers,
CC  including breast, ovarian, colon, lung and prostate cancer.
CC  Her-2/neu (see AAM51143) is a member of the tyrosine kinase family
CC  of receptor-like glycoproteins. It comprises an extracellular
CC  domain with homology to the epidermal growth factor receptor
CC  (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC  intracellular domain that also shows homology to EGFR. Its
CC  overexpression correlates with a poor prognosis in breast and
CC  ovarian cancers. The invention provides Her-2/neu fusion
CC  proteins, nucleic acids encoding them, viral vectors, and vaccines
CC  comprising the fusion proteins or nucleic acid molecules. In
CC  preferred fusion proteins, the extracellular domain of a Her-2/neu
CC  protein is fused to a Her-2/neu intracellular domain or
CC  phosphorylation domain (or its DeltaCD fragment). An immune
CC  response to Her-2/neu protein is elicited or enhanced by
CC  administering the fusion protein in the form of a vaccine, or by
CC  transfecting cells of an animal ex vivo with a nucleic acid
CC  encoding the fusion protein, and delivering the transfected cells
CC  to the animal. The fusion proteins, nucleic acids, and isolated
CC  specific T-cells are useful for inhibiting the development of a
CC  cancer, especially breast, ovarian, colon, lung or prostate cancer
CC  in a patient. T cells that specifically react with a Her-2/neu
CC  fusion protein can be used to remove tumour cells from a sample in
CC  order to inhibit the development of cancer in a patient.
XX
SQ  Sequence 653 AA:

Query Match      81.1%; Score 340; DB 23; Length 653;
Best Local Similarity 100.0%; Pred. No. 4e-302;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 MELAALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHYOGCQVVOGNTL 60
DB  1 MELAALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHYOGCQVVOGNTL 60

```

QY 61 ELTYLPTNALSFLQDIQEVGVYLIANQVPLQRLRIYRGTLFEDNYALAVLNG 120  
 |||||  
 Db 61 ELTYLPTNALSFLQDIQEVGVYLIANQVPLQRLRIYRGTLFEDNYALAVLNG 120  
 QY 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNQLA 180  
 |||||  
 Db 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNQLA 180  
 QY 181 LTIIDTNRSRACHPCSPMKGSKRCWGESSESDCOSLTRVCAGGCARCKGPLPTDCCHEOC 240  
 |||||  
 Db 181 LTIIDTNRSRACHPCSPMKGSKRCWGESSESDCOSLTRVCAGGCARCKGPLPTDCCHEOC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPEGRTYFGASCVTACP 300  
 |||||  
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPEGRTYFGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCERCKSPCAR 340  
 |||||  
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCERCKSPCAR 340

## RESULT 5

AAB21204

ID AAB21204 standard; protein; 712 AA.

AC AAB21204;

DT 12-JAN-2001 (first entry)

DE Human HER-2/neu fusion protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;

XX breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX colon cancer; fusion protein.

XX Homo sapiens.

OS Synthetic.

PN WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.

XX (SMK) SMITHKLINE BEECHAM.

XX Cheever MA, Gheysen D;

XX WPI: 2000-505976/45.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins

XX useful for vaccinating against breast, ovarian, colon, lung and

XX prostate cancers -

XX Claim 27; Fig 13; 128pp; English.

XX The present sequence is a fusion protein comprising the extracellular  
 CC domain and a preferred portion of the phosphorylation domain of the human  
 CC HER-2/neu protein. HER-2/neu is a member of the tyrosine kinase family of  
 CC receptor-like glycoproteins and shows homology to the epidermal growth  
 CC factor receptor (EGFR). It probably plays a part in cell growth and/or  
 CC differentiation. The HER-2/neu gene is an oncogene. HER-2/neu fusion  
 CC proteins may be used to treat or prevent cancer by eliciting or enhancing  
 CC an immune response to the HER-2/neu protein. They may be used to treat  
 CC malignancies such as breast, ovarian, colon, lung and prostate cancers,  
 CC and may be used as an antigen to vaccinate against these neoplasias.

PS Sequence 712 AA;

XX Query Match

81.1%; Score 340; DB 21; Length 712;

Best Local Similarity 100.0%; Pred. No. 4,3e-302;  
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAAACRWGLLALLPPEAASVQVCTGDMKRLPASPETHLMDIRHLXQCCQVQVGNL 60  
 |||||  
 Db 1 MELAAACRWGLLALLPPEAASVQVCTGDMKRLPASPETHLMDIRHLXQCCQVQVGNL 60  
 QY 61 ELTYLPTNALSFLQDIQEVGVYLIANQVPLQRLRIYRGTLFEDNYALAVLNG 120  
 |||||  
 Db 61 ELTYLPTNALSFLQDIQEVGVYLIANQVPLQRLRIYRGTLFEDNYALAVLNG 120  
 QY 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNQLA 180  
 |||||  
 Db 121 DPLNNTPTVGTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNQLA 180  
 QY 181 LTIIDTNRSRACHPCSPMKGSKRCWGESSESDCOSLTRVCAGGCARCKGPLPTDCCHEOC 240  
 |||||  
 Db 181 LTIIDTNRSRACHPCSPMKGSKRCWGESSESDCOSLTRVCAGGCARCKGPLPTDCCHEOC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPEGRTYFGASCVTACP 300  
 |||||  
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPEGRTYFGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCERCKSPCAR 340  
 |||||  
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCERCKSPCAR 340

## RESULT 6

AAM51149

ID AAM51149 standard; Protein; 712 AA.

AC AAM51149;

DT 17-JUN-2002 (first entry)

DE Her-2/neu extracellular domain-delta-phosphorylation domain fusion.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Location/Qualifiers

XX Key 1..653

XX Domain /note="extracellular domain"

XX Domain 654..712

XX FT /note="phosphorylation domain fragment"

XX WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI: 2002-241743/29.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting  
 PT or enhancing an immune response to the protein, has Her-2/neu  
 PT extracellular domain fused to Her-2/neu intracellular or  
 PT phosphorylation domain -

PS Claim 37; Fig 13; 141pp; English.

XX The present sequence is that of a fusion protein between the  
 CC extracellular domain and a fragment (Delta) of the phosphorylation

CC domain of human Her-2/neu (see AAM5143), an oncogenic self-protein  
CC and target for anti-cancer vaccines. The fusion protein can be  
CC obtained by recombinant DNA methods. Her-2/neu overexpression  
CC correlates with a poor prognosis in breast and ovarian cancers.  
CC The invention provides Her-2/neu fusion proteins, nucleic acids  
CC encoding them, viral vectors, and vaccines comprising the fusion  
CC proteins or nucleic acid molecules. In preferred fusion proteins,  
CC the extracellular domain of Her-2/neu is fused to a Her-2/neu  
CC intracellular domain or phosphorylation domain (or its Deltap  
CC fragment). An immune response to Her-2/neu protein is elicited or  
CC enhanced by administering the fusion protein in the form of a vaccine,  
CC or by transfecting cells of an animal ex vivo with a nucleic acid  
CC encoding the fusion protein, and delivering the transfected cells  
CC to the animal. The fusion proteins, nucleic acids, and isolated  
CC specific T-cells are useful for inhibiting the development of a  
CC cancer, especially breast, ovarian, colon, lung or prostate cancer  
CC in a patient. T cells that specifically react with a Her-2/neu  
CC fusion protein can be used to remove tumour cells from a sample in  
CC order to inhibit the development of cancer in a patient.

SQ Sequence 712 AA:

Query Match 81.1%; Score 340; DB 23; Length 712;  
Best Local Similarity 100.0%; Pred. No. 4.3e-302;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 METALACRMGLLLALLPRGAASQVCTGTDMLRLPASPEITHDMLRHLYOGCQVVGNL 60  
DB 1 METALACRMGLLLALLPRGAASQVCTGTDMLRLPASPEITHDMLRHLYOGCQVVGNL 60  
OY 61 ELTYLPTNASLSFLQDIOEVGYVLLAHNOVROVPLQRLRIVRGTOLEEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVGYVLLAHNOVROVPLQRLRIVRGTOLEEDNYALAVLDNG 120  
OY 121 DPLNNTPTVVGASPGGLREQLSLRTEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVVGASPGGLREQLSLRTEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180  
OY 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVAGGACARCKGLPTDCCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVAGGACARCKGLPTDCCHEQC 240  
OY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300  
OY 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKCSKPCAR 340  
DB 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKCSKPCAR 340

RESULT 7

ID AAM19764 standard; Protein: 782 AA.

XX AAM19764;

DT 17-SEP-1997 (first entry)

DE Her2-GM-CSF immunostimulant fusion protein.

XX Her2-GM-CSF; granulocyte macrophage colony stimulating factor;

KM growth factor receptor; oncogene; immunostimulant; cancer;

XX therapy.

OS Homo sapiens.

XX Key

FH Protein

FT Peptide

FT Protein

FT Protein

/label= GM-CSF

XX W09724438-A1.

XX 10-JUL-1997.

XX 23-DEC-1996; 96WO-US20241.

XX 28-DEC-1995; 95US-0579823.

XX (ACTI-) ACTIVATED CELL THERAPY INC.

XX Laus R, Ruegg CL, Wu H;

XX WPI: 1997-363674/33.

XX N-PSDB; AAT72725.

XX Potent APC that activates T-cells to give multivalent cellular

XX immune response - can also induce a cytotoxic T-cell response in a

XX vertebrate subject

XX Disclosure: Fig 8; 45pp; English.

XX A fusion protein (AAM19764) comprises Her2 (a growth factor receptor

XX that is overexpressed in breast, ovarian can other cancer cells)

XX and granulocyte-macrophage colony stimulating factor (GM-CSF). It

XX is the expression product of a nucleic acid molecule (AAT72725)

XX prep. by PCR amplification of Her2 cDNA from a breast cancer cell

XX line and fusion to GM-CSF cDNA. Fusion expression vectors can be

XX used to transfect mammalian and insect cells. The Her2-GM-CSF

XX fusion protein is used to generate anti-Her2 immunity. Tumour

XX cells are eliminated by cytotoxic T lymphocytes activated in vivo

XX or in vitro by exposure to antigen-presenting cells exposed to the

XX fusion protein.

SQ Sequence 782 AA:

Query Match 81.1%; Score 340; DB 18; Length 782;  
Best Local Similarity 100.0%; Pred. No. 4.6e-302;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 METALACRMGLLLALLPRGAASQVCTGTDMLRLPASPEITHDMLRHLYOGCQVVGNL 60  
DB 1 METALACRMGLLLALLPRGAASQVCTGTDMLRLPASPEITHDMLRHLYOGCQVVGNL 60  
OY 61 ELTYLPTNASLSFLQDIOEVGYVLLAHNOVROVPLQRLRIVRGTOLEEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVGYVLLAHNOVROVPLQRLRIVRGTOLEEDNYALAVLDNG 120  
OY 121 DPLNNTPTVVGASPGGLREQLSLRTEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVVGASPGGLREQLSLRTEILKGVLIQRNPOLCYODTILMKDIFHKNNOLA 180  
OY 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVAGGACARCKGLPTDCCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVAGGACARCKGLPTDCCHEQC 240  
OY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300  
OY 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKCSKPCAR 340  
DB 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKCSKPCAR 340

RESULT 8

ID AAB21203 standard; protein: 919 AA.

XX AAB21203;

DT 12-JAN-2001 (first entry)

```

XX Human HER-2/neu fusion protein.
DE
XX
XX Human: HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
KM breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer; fusion protein.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX WO200044899-A1.
PN
XX
XX 03-AUG-2000.
PD
XX
XX 28-JAN-2000; 2000WO-US02164.
PF
XX
XX 29-JAN-1999; 99US-0117976.
PR
XX
XX (CORI-) CORIAX CORP.
PA (SMIR ) SMITHKLINE BEECHAM.
PI
XX
XX Cheever MA, Gheysen D;
PI
XX WPI; 2000-505976/45.
DR
XX
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
XX Claim 2: Fig 12; 128pp; English.
PS
XX
XX The present sequence is a fusion protein comprising the extracellular
CC domain and the phosphorylation domain of the human HER-2/neu protein.
CC HER-2/neu is a member of the tyrosine kinase family of receptor-like
CC glycoproteins and shows homology to the epidermal growth factor receptor
CC (EGFR). It probably plays a part in cell growth and/or differentiation.
CC The HER-2/neu gene is an oncogene. HER-2/neu fusion proteins may be used
CC to treat or prevent cancer by eliciting or enhancing an immune response
CC to the HER-2/neu protein. They may be used to treat malignancies such as
CC breast, ovarian, colon, lung and prostate cancers, and may be used as an
CC antigen to vaccinate against these neoplasias.
XX
XX
SQ Sequence 919 AA:
Query Match 81.1%; Score 340; DB 21; length 919;
Best local Similarity 100.0%; Pred. No. 5.3e-302;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 METALCRMGLLLALLPPGAATQVCTGTDMKRLPASPETHLDMRLHLYOGCGVOGNTL 60
DB 1 METALCRMGLLLALLPPGAATQVCTGTDMKRLPASPETHLDMRLHLYOGCGVOGNTL 60
OY 61 ELTYLPNALSFLQDQEVGYVLIHNOVROYPLQRLKIVKRTQLFEDNYALAVDNG 120
DB 61 ELTYLPNALSFLQDQEVGYVLIHNOVROYPLQRLKIVKRTQLFEDNYALAVDNG 120
OY 121 DPLNTPVVGASPGGRLREVLQSLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNTPVVGASPGGRLREVLQSLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
OY 121 DPLNTPVVGASPGGRLREVLQSLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNTPVVGASPGGRLREVLQSLRSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDTNRSPACHPCSPMCKSGKRWGSSSDCSLRTVACGAGCARCKGPLPTDCHEQC 240
DB 181 LTLIDTNRSPACHPCSPMCKSGKRWGSSSDCSLRTVACGAGCARCKGPLPTDCHEQC 240
OY 241 AAGCTGRKHSDDLACHFNISGICELHCPALVYNTDFFESMPREKRYTFGASCVTACP 300
DB 241 AAGCTGRKHSDDLACHFNISGICELHCPALVYNTDFFESMPREKRYTFGASCVTACP 300
OY 301 YNVLSTDVGSCTLYCPLHNOEVTAEADGTORCEKSKPCAR 340
DB 301 YNVLSTDVGSCTLYCPLHNOEVTAEADGTORCEKSKPCAR 340

```

```

RESULT 9
ID AAM51148
XX AAM51148 standard; protein; 919 AA.
XX
XX AAM51148;
AC
XX
XX 17-JUN-2002 (first entry)
DT
XX
XX Her-2/neu extracellular domain-phosphorylation domain fusion.
DE
XX
XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Domain 1..653
FT Domain /note="extracellular domain"
FT Domain 654..919
FT Domain /note="phosphorylation domain"
XX
XX WO200212341-A2.
PN
XX
XX 14-FEB-2002.
PD
XX
XX 03-AUG-2001; 2001WO-US24283.
PF
XX
XX 03-AUG-2000; 2000US-0632507.
PR
XX
XX (CORI-) CORIAX CORP.
PA (SMIR ) SMITHKLINE BEECHAM BIOLOGICALS.
PI
XX
XX Cheever MA, Gheysen D;
PI
XX WPI; 2002-241743/29.
DR
XX
XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain -
XX
XX
XX Claim 2: Fig 12; 141pp; English.
PS
XX
XX The present sequence is that of a fusion protein between the
CC extracellular domain and phosphorylation domain of human Her-2/neu
CC (see AAM51143), an oncogenic self-protein and target for anti-cancer
CC vaccines. The fusion protein can be obtained by recombinant DNA
CC methods. Her-2/neu overexpression correlates with a poor prognosis
CC in breast and ovarian cancers. The invention provides Her-2/neu
CC fusion proteins, nucleic acids encoding them, viral vectors, and
CC vaccines comprising the fusion proteins or nucleic acid molecules.
CC In preferred fusion proteins, the extracellular domain of a
CC Her-2/neu protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaCD fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfecting cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.
XX
XX
SQ Sequence 919 AA:
Query Match 81.1%; Score 340; DB 23; length 919;
Best local Similarity 100.0%; Pred. No. 5.3e-302;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 METALCRMGLLLALLPPGAATQVCTGTDMKRLPASPETHLDMRLHLYOGCGVOGNTL 60
DB 1 METALCRMGLLLALLPPGAATQVCTGTDMKRLPASPETHLDMRLHLYOGCGVOGNTL 60

```

```

Db      1  MEALALCRWGLLLALLPPGAASQVCTGTDMKLRLPASPTHLDMRLHLYOGCQVVGNL 60
QY      61  ELTYLPTNASLSLFODIQEOGVYLIANOVROYPLQRLRIVRGTOLEFEDNYALAVDNG 120
      |||
      61  ELTYLPTNASLSLFODIQEOGVYLIANOVROYPLQRLRIVRGTOLEFEDNYALAVDNG 120
QY      121  DPLNNTPTVVGASPGGLREQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
      |||
      121  DPLNNTPTVVGASPGGLREQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
Db      121  DPLNNTPTVVGASPGGLREQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
QY      181  LTLIDTNRSRACHPCSPMCKGSRGWGESSDQSLRTVCAGGCARCKGPLPTDCCHEQC 240
      |||
      181  LTLIDTNRSRACHPCSPMCKGSRGWGESSDQSLRTVCAGGCARCKGPLPTDCCHEQC 240
Db      181  LTLIDTNRSRACHPCSPMCKGSRGWGESSDQSLRTVCAGGCARCKGPLPTDCCHEQC 240
QY      241  AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
      |||
      241  AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
Db      241  AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
QY      301  YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCAR 340
      |||
      301  YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCAR 340
Db      301  YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCAR 340

RESULT 10
AAB21208
ID      AAB21208 standard; Protein; 1200 AA.
XX
AC      AAB21208;
XX
DT      12-JAN-2001 (first entry)
XX
DE      Human HER-2/neu protein.
XX
KM      Human: HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KM      breast cancer; prostate cancer; ovarian cancer; lung cancer;
KM      colon cancer.
XX
OS      Homo sapiens.
XX
PN      WO200044899-A1.
XX
PD      03-AUG-2000.
XX
PF      28-JAN-2000; 2000WO-US02164.
XX
PR      29-JAN-1999; 99US-0117976.
XX
PA      (CORI-) CORIXA CORP.
PA      (SMIK-) SMITHKLINE BEECHAM.
XX
PI      Cheever MA, Gheysen D;
XX
DR      WPI; 2000-505976/45.
DR      N-PSDB; AAA89736.
XX
PT      HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT      useful for vaccinating against breast, ovarian, colon, lung and
PT      prostate cancers -
XX
PS      disclosure; Fig 15; 128pp; English.
XX
CC      The present sequence is the human HER-2/neu protein. It is a member
CC      of the tyrosine kinase family of receptor-like glycoproteins and shows
CC      homology to the epidermal growth factor receptor (EGFR). It probably
CC      plays a part in cell growth and/or differentiation. The HER-2/neu
CC      gene is an oncogene. An HER-2/neu fusion protein comprising a
CC      HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC      domain may be used to treat or prevent cancer by eliciting or
CC      enhancing an immune response to the HER-2/neu protein. It may be used
CC      to treat malignancies such as breast, ovarian, colon, lung and
CC      prostate cancers, and may be used as an antigen to vaccinate against
CC      these neoplasias.
XX

```

```

SQ      Sequence      1200 AA:
Query Match      81.1%; Score 340; DB 21; Length 1200;
Best Local Similarity 100.0%; Pred. No. 6.6e-302;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MEALALCRWGLLLALLPPGAASQVCTGTDMKLRLPASPTHLDMRLHLYOGCQVVGNL 60
      |||
      1  MEALALCRWGLLLALLPPGAASQVCTGTDMKLRLPASPTHLDMRLHLYOGCQVVGNL 60
Db      1  MEALALCRWGLLLALLPPGAASQVCTGTDMKLRLPASPTHLDMRLHLYOGCQVVGNL 60
QY      61  ELTYLPTNASLSLFODIQEOGVYLIANOVROYPLQRLRIVRGTOLEFEDNYALAVDNG 120
      |||
      61  ELTYLPTNASLSLFODIQEOGVYLIANOVROYPLQRLRIVRGTOLEFEDNYALAVDNG 120
Db      61  ELTYLPTNASLSLFODIQEOGVYLIANOVROYPLQRLRIVRGTOLEFEDNYALAVDNG 120
QY      121  DPLNNTPTVVGASPGGLREQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
      |||
      121  DPLNNTPTVVGASPGGLREQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
Db      121  DPLNNTPTVVGASPGGLREQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
QY      181  LTLIDTNRSRACHPCSPMCKGSRGWGESSDQSLRTVCAGGCARCKGPLPTDCCHEQC 240
      |||
      181  LTLIDTNRSRACHPCSPMCKGSRGWGESSDQSLRTVCAGGCARCKGPLPTDCCHEQC 240
Db      181  LTLIDTNRSRACHPCSPMCKGSRGWGESSDQSLRTVCAGGCARCKGPLPTDCCHEQC 240
QY      241  AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
      |||
      241  AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
Db      241  AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
QY      301  YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCAR 340
      |||
      301  YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCAR 340
Db      301  YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCAR 340

RESULT 11
AAW01111
ID      AAW01111 standard; Protein; 1255 AA.
XX
AC      AAW01111;
XX
DT      01-JAN-1997 (first entry)
XX
DE      HER-2/neu protein.
XX
KM      HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KM      breast cancer; ovary cancer; colon cancer; lung cancer;
KM      prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS      Homo sapiens.
XX
PN      WO9630514-A1.
XX
PD      03-OCT-1996.
XX
PF      28-MAR-1996; 96WO-US01689.
XX
PR      31-MAR-1995; 95US-0414417.
XX
PA      (UNIW) UNIV WASHINGTON.
XX
PI      Cheever MA, Disis ML;
XX
DR      WPI; 1996-455361/45.
DR      N-PSDB; AAT40739.
XX
PT      DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT      treatment of malignancies with which the HER-2/neu oncogene is
PT      associated
XX
PS      Claim 2; Page 56-61; 71pp; English.
XX

```

CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is  
CC the product of the HER-2/neu oncogene (see also AA140739). The  
CC protein is over-expressed in various cancers, including breast,  
CC ovarian, colon, lung and prostate. The intracellular domain of the  
CC protein can be used to immunise an animal against a malignancy with  
CC which the oncogene is associated. The polypeptide can be produced  
CC in transformed host cells for use in immunisation. Alternatively,  
CC animal cells are transfected in vivo or ex vivo with a viral vector  
CC that directs expression of the polypeptide.

XX  
SQ Sequence 1255 AA:

Query Match 81.1%; Score 340; DB 17; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 6.9e-302;

Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRMGGLLALLPPGAASVQVCTGTDMLRLPASPEHIDMLRHLYQGQVVGNL 60  
DB 1 MELAALCRMGGLLALLPPGAASVQVCTGTDMLRLPASPEHIDMLRHLYQGQVVGNL 60

QY 61 ELTYLPTNASLFLQDIOEVQGYVLIANQVROVPLQRLRIVGTQLFEDNVALAVLDNG 120  
DB 61 ELTYLPTNASLFLQDIOEVQGYVLIANQVROVPLQRLRIVGTQLFEDNVALAVLDNG 120

QY 121 DPLNNTTPVVGASPGGLRELQLSLFEILKGVLIQNRNPOLCYODITLMDIFHKNNOLA 180  
DB 121 DPLNNTTPVVGASPGGLRELQLSLFEILKGVLIQNRNPOLCYODITLMDIFHKNNOLA 180

QY 181 LTLIDTNRSRACHPCSPMCKSGSRGWESSSDQSLRTVYVAGGCAKCGPLPTDCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPMCKSGSRGWESSSDQSLRTVYVAGGCAKCGPLPTDCHEQC 240

QY 241 AACGTGPKHSDCLACLFHNSGICELHCPALVYNTDFFESMPNPEGRTYFGASCVTACP 300  
DB 241 AACGTGPKHSDCLACLFHNSGICELHCPALVYNTDFFESMPNPEGRTYFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCAR 340  
DB 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCAR 340

RESULT 12  
ID AAW92406 standard; Protein; 1255 AA.

XX AAW92406;

XX 21-APR-1999 (first entry)

XX Human HER-2/neu oncogene protein.

XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
XX malignancy; treatment; tumour.

OS Homo sapiens.

XX Key Location/Qualifiers  
XX Region 676..1255

XX /note="region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 9605-0625101.

XX 01-APR-1996; 9605-0625101.

XX 17-MAR-1993; 9305-0033644.

XX 12-AUG-1993; 9305-0106112.

XX 31-MAR-1995; 9505-0414417.

XX (UNIW ) UNIV WASHINGTON.

PI Cheever MA, Disis ML;

XX WPI; 1999-152835/13.

XX N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to  
XX an HER-2/neu associated malignancy, particularly for treating or  
XX preventing tumours

PS Claim 3; Column 31-38; 26pp; English.

XX This sequence represents the human HER-2/neu oncogene protein. A fragment  
XX of this protein is used in a method for eliciting or enhancing an immune  
XX response to HER-2/neu protein. The polypeptide can stimulate T cells and  
XX B cells to produce an immune response to the HER-2/neu protein. The  
XX method can be used for immunisation against a malignancy in which the  
XX HER-2/neu oncogene is associated and in the treatment of an existing  
XX tumour, or to prevent tumour occurrence or recurrence.

XX  
SQ Sequence 1255 AA:

Query Match 81.1%; Score 340; DB 20; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 6.9e-302;

Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRMGGLLALLPPGAASVQVCTGTDMLRLPASPEHIDMLRHLYQGQVVGNL 60  
DB 1 MELAALCRMGGLLALLPPGAASVQVCTGTDMLRLPASPEHIDMLRHLYQGQVVGNL 60

QY 61 ELTYLPTNASLFLQDIOEVQGYVLIANQVROVPLQRLRIVGTQLFEDNVALAVLDNG 120  
DB 61 ELTYLPTNASLFLQDIOEVQGYVLIANQVROVPLQRLRIVGTQLFEDNVALAVLDNG 120

QY 121 DPLNNTTPVVGASPGGLRELQLSLFEILKGVLIQNRNPOLCYODITLMDIFHKNNOLA 180  
DB 121 DPLNNTTPVVGASPGGLRELQLSLFEILKGVLIQNRNPOLCYODITLMDIFHKNNOLA 180

QY 181 LTLIDTNRSRACHPCSPMCKSGSRGWESSSDQSLRTVYVAGGCAKCGPLPTDCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPMCKSGSRGWESSSDQSLRTVYVAGGCAKCGPLPTDCHEQC 240

QY 241 AACGTGPKHSDCLACLFHNSGICELHCPALVYNTDFFESMPNPEGRTYFGASCVTACP 300  
DB 241 AACGTGPKHSDCLACLFHNSGICELHCPALVYNTDFFESMPNPEGRTYFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCAR 340  
DB 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCAR 340

RESULT 13  
ID AAB21198 standard; protein; 1255 AA.

XX AAB21198;

XX 12-JAN-2001 (first entry)

XX Human HER-2/neu protein..

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;  
XX breast cancer; prostate cancer; ovarian cancer; lung cancer;  
XX colon cancer.

XX Homo sapiens.

XX WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 9905-0117976.

```

XX (CORI-) CORIXA CORP.
PA (SMIK ) SMITHKLINE BEECHAM.
XX
XX Cheever MA, Gheysen D;
XX
XX WPI: 2000-505976/45.
XX N-PSDB: AAA89736.
XX
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
XX useful for vaccinating against breast, ovarian, colon, lung and
XX prostate cancers -
XX
XX Claim 52; Fig 7; 128pp; English.
XX
XX The present sequence is the human HER-2/neu protein. It is a member of
XX the tyrosine kinase family of receptor-like glycoproteins and shows
XX homology to the epidermal growth factor receptor (EGFR). It probably
XX plays a part in cell growth and/or differentiation. The HER-2/neu
XX gene is an oncogene. An HER-2/neu fusion protein comprising a
XX HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
XX domain may be used to treat or prevent cancer by eliciting or
XX enhancing an immune response to the HER-2/neu protein. It may be used
XX to treat malignancies such as breast, ovarian, colon, lung and
XX prostate cancers, and may be used as an antigen to vaccinate against
XX these neoplasias.
XX
XX Sequence 1255 AA:
SQ
XX
XX Query Match 81.1%; Score 340; DB 21; Length 1255;
XX Best Local Similarity 100.0%; Pred. No. 6.9e-302;
XX Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MELAALCRNGLLALLPPAASVQVCTGDMKRLRPASPTHLDMLRHLXGCGVQVGNL 60
XX |
XX | 1 MELAALCRNGLLALLPPAASVQVCTGDMKRLRPASPTHLDMLRHLXGCGVQVGNL 60
XX |
XX
XX QY 61 ELTYLPTNASLFLQDIOEVQGYVLLAHNOVRQVPLQRLRIYVGTQLFEDNVALAVLDNG 120
XX |
XX | 61 ELTYLPTNASLFLQDIOEVQGYVLLAHNOVRQVPLQRLRIYVGTQLFEDNVALAVLDNG 120
XX |
XX
XX DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNOLA 180
XX |
XX | 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNOLA 180
XX |
XX
XX DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNOLA 180
XX |
XX | 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNOLA 180
XX |
XX
XX QY 181 LFLIDNRRACHPCSPMKCKSGKMGESSEDCQSLTRTYCAGGACARCKPRLPDCCHEQC 240
XX |
XX | 181 LFLIDNRRACHPCSPMKCKSGKMGESSEDCQSLTRTYCAGGACARCKPRLPDCCHEQC 240
XX |
XX
XX DB 181 LFLIDNRRACHPCSPMKCKSGKMGESSEDCQSLTRTYCAGGACARCKPRLPDCCHEQC 240
XX |
XX | 181 LFLIDNRRACHPCSPMKCKSGKMGESSEDCQSLTRTYCAGGACARCKPRLPDCCHEQC 240
XX |
XX
XX QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP 300
XX |
XX | 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP 300
XX |
XX
XX DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP 300
XX |
XX | 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP 300
XX |
XX
XX QY 301 YNVLSTDVGSCTLVCPHLHNOEVTAEEDGTORCEKCSKPCAR 340
XX |
XX | 301 YNVLSTDVGSCTLVCPHLHNOEVTAEEDGTORCEKCSKPCAR 340
XX |
XX
XX DB 301 YNVLSTDVGSCTLVCPHLHNOEVTAEEDGTORCEKCSKPCAR 340
XX |
XX | 301 YNVLSTDVGSCTLVCPHLHNOEVTAEEDGTORCEKCSKPCAR 340
XX |
XX
XX RESULT 14
XX AA84780
XX ID AA84780 standard; Protein; 1255 AA.
XX
XX AC AA84780;
XX
XX DT 08-AUG-2000 (first entry)
XX
XX DE Amino acid sequence of the SPLICE erB-2 receptor protein.
XX
XX KW SPLICE erB-2 receptor protein; cell transformation disorder; cancer;
XX tumor cell proliferation; tissue degeneration; arthropathy;
XX bone resorption; inflammatory disease; degenerative disorder;
XX wound healing.
XX

```

```

OS Homo sapiens.
XX
XX MO200020579-A1.
XX
XX 13-APR-2000.
XX
XX PF 01-OCT-1999; 99WO-CA00912.
XX
XX PR 02-OCT-1998; 98US-0165192.
XX
XX (UYMC-) UNIV MCMASSTER.
XX
XX Muller WJ, Siegel PM;
XX
XX WPI: 2000-303768/26.
XX N-PSDB: AAA14812.
XX
XX Nucleic acid encoding an erB 2 receptor protein designated SPLICE
XX erB-2, inhibitors of the protein are useful for treatment of cancer -
XX
XX Claim 3; Fig 2; 60pp; English.
XX
XX The present sequence represents a SPLICE erB-2 receptor protein. The
XX protein has an in-frame deletion of 16 amino acids, 2 of which are
XX conserved cysteine residues, compared to the unspliced protein. The
XX erB-2 polynucleotide is used to construct probes for detecting
XX disorders of cell transformation such as cancer. Antipodes to the
XX protein may be used to detect SPLICE erB-2 in a sample. Agents
XX (e.g. antisense oligonucleotides) which inhibit the expression of
XX SPLICE erB-2 are useful for reducing tumor cell proliferation and
XX treating cancer. Substances which stimulate SPLICE erB-2 are useful
XX for treating conditions involving damaged cells including conditions
XX in which degeneration of tissue occurs, such as arthropathy, bone
XX resorption, inflammatory diseases, degenerative disorders of the
XX central nervous system and wound healing.
XX
XX Sequence 1255 AA:
SQ
XX
XX Query Match 81.1%; Score 340; DB 21; Length 1255;
XX Best Local Similarity 100.0%; Pred. No. 6.9e-302;
XX Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MELAALCRNGLLALLPPAASVQVCTGDMKRLRPASPTHLDMLRHLXGCGVQVGNL 60
XX |
XX | 1 MELAALCRNGLLALLPPAASVQVCTGDMKRLRPASPTHLDMLRHLXGCGVQVGNL 60
XX |
XX
XX DB 1 MELAALCRNGLLALLPPAASVQVCTGDMKRLRPASPTHLDMLRHLXGCGVQVGNL 60
XX |
XX | 1 MELAALCRNGLLALLPPAASVQVCTGDMKRLRPASPTHLDMLRHLXGCGVQVGNL 60
XX |
XX
XX QY 61 ELTYLPTNASLFLQDIOEVQGYVLLAHNOVRQVPLQRLRIYVGTQLFEDNVALAVLDNG 120
XX |
XX | 61 ELTYLPTNASLFLQDIOEVQGYVLLAHNOVRQVPLQRLRIYVGTQLFEDNVALAVLDNG 120
XX |
XX
XX DB 61 ELTYLPTNASLFLQDIOEVQGYVLLAHNOVRQVPLQRLRIYVGTQLFEDNVALAVLDNG 120
XX |
XX | 61 ELTYLPTNASLFLQDIOEVQGYVLLAHNOVRQVPLQRLRIYVGTQLFEDNVALAVLDNG 120
XX |
XX
XX QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNOLA 180
XX |
XX | 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNOLA 180
XX |
XX
XX DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNOLA 180
XX |
XX | 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNOLA 180
XX |
XX
XX QY 181 LFLIDNRRACHPCSPMKCKSGKMGESSEDCQSLTRTYCAGGACARCKPRLPDCCHEQC 240
XX |
XX | 181 LFLIDNRRACHPCSPMKCKSGKMGESSEDCQSLTRTYCAGGACARCKPRLPDCCHEQC 240
XX |
XX
XX DB 181 LFLIDNRRACHPCSPMKCKSGKMGESSEDCQSLTRTYCAGGACARCKPRLPDCCHEQC 240
XX |
XX | 181 LFLIDNRRACHPCSPMKCKSGKMGESSEDCQSLTRTYCAGGACARCKPRLPDCCHEQC 240
XX |
XX
XX QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP 300
XX |
XX | 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP 300
XX |
XX
XX DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP 300
XX |
XX | 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCVTACP 300
XX |
XX
XX QY 301 YNVLSTDVGSCTLVCPHLHNOEVTAEEDGTORCEKCSKPCAR 340
XX |
XX | 301 YNVLSTDVGSCTLVCPHLHNOEVTAEEDGTORCEKCSKPCAR 340
XX |
XX
XX DB 301 YNVLSTDVGSCTLVCPHLHNOEVTAEEDGTORCEKCSKPCAR 340
XX |
XX | 301 YNVLSTDVGSCTLVCPHLHNOEVTAEEDGTORCEKCSKPCAR 340
XX |
XX
XX RESULT 15
XX AA92620
XX ID AA92620 standard; Protein; 1255 AA.
XX
XX AC AA92620;
XX
XX

```

DT 10-AUG-2000 (first entry)  
XX  
XX Human heregulin 2 (Her2).  
XX  
XX Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;  
KW self-protein; cancer; breast cancer; prostate cancer;  
XX cell-associated peptide antigen; foreign epitope.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FH Domain  
FT Location/Qualifiers  
FT 1..173  
FT /label= N-terminal  
FT /note= "mature polypeptide"  
FT 5..25  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 59..73  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 103..117  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 149..163  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 174..323  
FT /label= Cysteine\_rich\_domain  
FT 210..224  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 250..264  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 324..483  
FT /label= Ligand\_binding\_domain  
FT 325..339  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 369..383  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 465..479  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 484..623  
FT /label= Cysteine\_rich\_domain  
FT 579..593  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 624..654  
FT /label= Transmembrane\_domain  
FT 632..652  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 653..667  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 655..1010  
FT /label= Tyrosine\_kinase\_domain  
FT 661..675  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 695..709  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 710..730  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 1011..1235  
FT /label= C-terminal\_domain  
XX  
PN WO200020027-A2.

XX 13-APR-2000.  
PD  
XX 05-OCT-1999; 99WO-DK00525.  
XX  
XX 05-OCT-1998; 98DK-0001261.  
PR 20-OCT-1998; 98US-0105011.  
XX  
XX (MERI-) M & E BIOTECH AS.  
XX  
PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;  
PI Gautam A, Birk P, Karlsson G;  
XX  
XX WPI: 2000-349917/30.  
DR N-PSDB; AAA09455.  
XX  
XX Inducing immune responses to weakly immunogenic, tumor associated  
PT peptide antigens for the treatment of breast and prostate cancer  
XX  
XX Claim 62; Page 193-198; 220pp; English.  
XX  
CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
CC Her2 can be used in the claimed method as an autovaccine to induce a CTL  
CC response. Subdominant CTL epitopes, antibody binding regions and  
CC cysteine residues involved in disulfide bonds are preserved in the  
CC immunogenized forms. Regions suitable for the insertion of foreign T  
CC helper epitopes were identified (see features table). The method  
CC is used for inducing immune responses against weakly immunogenic  
CC cell-associated peptide antigens (PA) such as those associated with  
CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
CC The method comprises effecting simultaneous presentation by antigen  
CC producing cells (APCs) of the animals immune system of: (1) at least 1  
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
CC B-cell group derived from the cell-associated PA; and (2) at least 1  
CC first T helper cell group which is foreign to the animal. Analogues of  
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
CC part of all known and predicted CTL and B-cell epitopes of the respective  
CC PA and including at least one foreign T helper epitope are also claimed.  
CC The method is used to treat prostate, prostate/breast or breast cancer  
CC when the PA is human PSM, FGF8b and Her2, respectively.  
XX  
SQ Sequence 1255 AA;  
Query Match 81.1%; Score 340; DB 21; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 6; 9e-302;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPPGAATQVCTGTDMLRLPASETHLDMRLHYGCGVVGNL 60  
DB 1 MELAALCRWGLLLALLPPGAATQVCTGTDMLRLPASETHLDMRLHYGCGVVGNL 60  
QY 61 ELTYLEPTNASLSTFLDIOEVQGVYLLAHNOVQVPLQRLIRYRGQLFEDNALAVLDNG 120  
DB 61 ELTYLEPTNASLSTFLDIOEVQGVYLLAHNOVQVPLQRLIRYRGQLFEDNALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGCLRELOLSLFEILKGVLIORNPOLCYQDTILKKDIFHKNNOLA 180  
DB 121 DPLNNTPTVTGASPGCLRELOLSLFEILKGVLIORNPOLCYQDTILKKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMKGSRKMGSESDCSLRTVACAGCARKGRLPTDCCHQC 240  
DB 181 LTLIDNRSRACHPCSPMKGSRKMGSESDCSLRTVACAGCARKGRLPTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTRESMPNPEGRTFFGASCTYACP 300  
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTRESMPNPEGRTFFGASCTYACP 300  
QY 301 YNTLSTDVGSCTLYVCPLHNOEYTAEDGTORCKGSPCAR 340  
DB 301 YNTLSTDVGSCTLYVCPLHNOEYTAEDGTORCKGSPCAR 340



Search completed: March 4, 2003, 12:44:57  
job time : 57.8474 secs

---



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:43:27 ; Search time 47.1165 seconds  
(without alignments)  
854,910 Million cell updates/sec

Title: US-09-234-208b-2  
Perfect score: 419  
Sequence: 1 MELAALCRWGLLALLPFGA.....VGRGPDPAHVAVNLSEYEG 419

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	340	81.1	1255	1 A24571	protein-tyrosine k
2	67	16.0	1254	2 I48161	p-185 precursor
3	52	12.4	1260	1 TVRNU	protein-tyrosine k
4	9	2.1	230	2 B82252	knife-related prote
5	9	2.1	527	2 A42032	epidermal growth f
6	9	2.1	644	2 A36325	epidermal growth f
7	9	2.1	1210	1 G0HUE	epidermal growth f
8	9	2.1	1210	2 A53183	epidermal growth f
9	9	2.1	1223	1 TVCHLV	epidermal growth f
10	8	1.9	277	2 JCS284	carboxyl reductase
11	8	1.9	341	2 B83298	conserved hypotet
12	8	1.9	348	2 D95067	phenylalanine-trna
13	8	1.9	375	2 C97935	phenylalanine-trna
14	8	1.9	463	2 H70922	hypothetical prote
15	8	1.9	520	2 C70311	hypothetical prote
16	8	1.9	583	2 D90052	hypothetical prote
17	8	1.9	725	2 IJMSNG	hypothetical prote
18	8	1.9	858	1 IJRTNC	neural cell adhesi
19	8	1.9	865	2 A85032	neural cell adhesi
20	8	1.9	879	2 S43910	hypothetical prote
21	8	1.9	1115	1 IJMSNL	chloroplast outer
22	8	1.9	1503	2 T01098	neural cell adhesi
23	7	1.7	62	2 F90177	chloroplast outer
24	7	1.7	104	2 G69482	hypothetical prote
25	7	1.7	111	2 C35826	hypothetical prote
26	7	1.7	121	2 AH1848	hypothetical 13k p
27	7	1.7	173	2 S10199	hypothetical prote
28	7	1.7	189	2 A70186	NADH2 dehydrogenas
29	7	1.7	189	2 S04670	neutrophil activat
					hypothetical prote

30	7	1.7	207	2 A69941	capsular polysacch
31	7	1.7	209	2 T00733	hypothetical prote
32	7	1.7	220	2 S05595	trypsin inhibitor
33	7	1.7	220	2 I50588	fibroblast growth
34	7	1.7	221	2 UC7587	stromal cell-deriv
35	7	1.7	223	2 H83462	heme exporter prot
36	7	1.7	224	1 A33861	trans-activating t
37	7	1.7	248	2 C83431	type III export pr
38	7	1.7	255	2 AH2460	glycosyltransferas
39	7	1.7	259	2 D75275	endonuclease III -
40	7	1.7	268	2 C70620	hypothetical prote
41	7	1.7	268	2 C90709	thiol disulfide in
42	7	1.7	268	2 G85559	thiol disulfide in
43	7	1.7	268	2 B64794	hypothetical prote
44	7	1.7	269	1 D43706	serine O-acetyltra
45	7	1.7	271	2 A45606	DNA-binding protei

ALIGNMENTS

RESULT 1  
A24571  
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human  
N:Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein e  
C:Species: Homo sapiens (man)  
C>Date: 25-Oct-1987 #sequence\_revision 06-Dec-1996 #text\_change 11-Jun-1999  
C:Accession: A24571, A25491, A44188, B44188, I59509, I57622  
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T  
Nature 319, 230-234, 1986  
A>Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growt  
A:Reference number: A24571; MUID:86118663; PMID:3003577  
A:Accession: A24571  
A:Molecule type: mRNA  
A:Residues: 1-1295 <YAM>  
A:Cross-references: GB:X03363; NID:931197; PIDN:CAA27060.1; PID:g31198  
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985  
A>Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid  
A:Reference number: A25491; MUID:86016729; PMID:2995967  
A:Accession: A25491  
A:Molecule type: DNA  
A:Residues: 737-1031 <SEM>  
A:Cross-references: GB:M11767; NID:9182163; PIDN:AAA35808.1; PID:g553282  
R:Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg  
Science 230, 1132-1139, 1985  
A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro  
A:Reference number: A44188; MUID:86070181; PMID:2999974  
A:Accession: A44188  
A:Molecule type: DNA  
A:Residues: 740-910 <COU1>  
A:Cross-references: GB:M12036; NID:9183988; PIDN:AAA35978.1; PID:g183989  
A:Accession: B44188  
A:Molecule type: mRNA  
A:Residues: 1-517, 'RALF', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>  
A:Cross-references: GB:M1730; NID:g183986  
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.  
Science 229, 974-976, 1985  
A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.  
A:Reference number: I59509; MUID:85272597; PMID:2992089  
A:Accession: I59509  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 832-909 <REX>  
A:Cross-references: GB:L29395; NID:9459807; PIDN:AAA35809.1; PID:g459808  
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.  
Mol. Cell. Biol. 7, 2597-2601, 1987  
A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcrip  
A:Reference number: I57622; MUID:87286898; PMID:3039551  
A:Accession: I57622  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:g183983; PIDN:AA58637.1; PID:g553332  
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
C:Genetics:  
A:Gene: GDB:ERBB2; NGD: NEU; HER-2  
A:Cross-references: GDB:120613; OMIM:164870  
A:Map position: 17q21.1-17q21.1  
A:Introns: 25/1; 75/3; 147/1; 883/3  
A>Note: the 1st of introns is incomplete  
C:Function:  
A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph  
inase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-125/Domain: protein-tyrosine kinase erbb2 #status predicted <MAT>  
F:22-653/Domain: extracellular #status predicted <EXT>  
F:70-304/Domain: EGF receptor extracellular domain repeat <EEL>  
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>  
F:654-675/Domain: transmembrane #status predicted <TM>  
F:676-1255/Domain: intracellular #status predicted <INT>  
F:718-983/Domain: protein kinase homology <KIN>  
F:726-734/Region: protein kinase ATP-binding motif  
F:68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict  
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
F:753/Active site: Lys #status predicted  
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 81.1%; Score 340; DB 1; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METALACRMGLLALLPFGAASVOTCTGDMKRLPASPETHLMDLRLHYOGCGVOGNL 60  
Db 1 METALACRMGLLALLPFGAASVOTCTGDMKRLPASPETHLMDLRLHYOGCGVOGNL 60

QY 61 ELTYLPNNAISLFQDIOEVGYLAAHNOYRQYPLDRLKIVKSTOLFEDNYALAVLDNG 120  
Db 61 ELTYLPNNAISLFQDIOEVGYLAAHNOYRQYPLDRLKIVKSTOLFEDNYALAVLDNG 120

QY 121 DPLNNTPTVVGASPGSGREVLRLSLTEILKGVLIQRNPOLCYODTILMDIFHKNNOLA 180  
Db 121 DPLNNTPTVVGASPGSGREVLRLSLTEILKGVLIQRNPOLCYODTILMDIFHKNNOLA 180

QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLRTVACGAGCARCKGLPTDCHEQC 240  
Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLRTVACGAGCARCKGLPTDCHEQC 240

QY 241 AAGCTGKHSDDCLACLFHNSGICELHCPALVYNTDTFESMPREGRYTFGASCVTACP 300  
Db 241 AAGCTGKHSDDCLACLFHNSGICELHCPALVYNTDTFESMPREGRYTFGASCVTACP 300

QY 301 YNYLSTDVSGCTLVCPAHNOEVTAEADTGRCCKSPCAR 340  
Db 301 YNYLSTDVSGCTLVCPAHNOEVTAEADTGRCCKSPCAR 340

RESULT 2  
148161  
P:185 precursor - golden hamster  
C:Species: Mesocricetus auratus (golden hamster)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
C:Accession: 148161  
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,  
Gene 140, 251-255, 1994  
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
A:Reference number: 148161; MIMD:94193007; PMID:7908275  
A:Accession: 148161  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1254 <RES>  
A:Cross-references: GB:D16295; NID:g493236; PIDN:BA03801.1; PID:g747595  
C:Genetics:  
A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP  
F:718-983/Domain: protein kinase homology <KIN>  
F:726-734/Region: protein kinase ATP-binding motif

Query Match 16.0%; Score 67; DB 2; Length 1254;  
Best Local Similarity 100.0%; Pred. No. 7,4e-58;  
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 LPTDCHGQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPREGRYT 290  
Db 231 LPTDCHGQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPREGRYT 290

QY 291 FGASCVT 297  
Db 291 FGASCVT 297

RESULT 3  
TVRTNU  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 11-Jun-1999  
C:Accession: A24562; A61204  
R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.  
Nature 319, 226-230, 1986  
A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein  
A:Reference number: A24562; MIMD:86118662; PMID:3945311  
A:Accession: A24562  
A:Molecule type: mRNA  
A:Residues: 1-1260 <BAR>  
A:Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746  
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohe  
Carcinogenesis 12, 1975-1978, 1991  
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals n  
2-thiazolyl]formamide or N-methyl-N-nitrosourea.  
A:Reference number: A61204; MIMD:92035293; PMID:1682063  
A:Accession: A61204  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 637-663, 'V', 665-702 <MAS>  
A>Note: authors translated the codon GCA for residue 25 as Val  
C:Genetics:  
A:Gene: neu  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1260/Domain: protein-tyrosine kinase neu #status predicted <MAT>  
F:658-680/Domain: transmembrane #status predicted <TM>  
F:723-988/Domain: protein kinase homology <KIN>  
F:731-739/Region: protein kinase ATP-binding motif  
F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicte  
F:691/Binding site: phosphate (Thr) (covalent) #status predicted  
F:758/Active site: Lys #status predicted  
F:882,1227,1233/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 12.4%; Score 52; DB 1; Length 1260;  
Best Local Similarity 100.0%; Pred. No. 6,2e-43;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 LPTDCHGQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESM 282  
Db 235 LPTDCHGQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESM 286

RESULT 4  
B82252  
R:ite-related protein VC01012 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C:Species: Vibrio cholerae  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: B82252  
R:Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers

L.R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 A:Reference number: A82035; MUID:20406833; PMID:10952301  
 A:Accession: B82252  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-230 <HE>  
 A:Cross-references: GB:AE004183; GB:AE003852; NID:g9655473; PIDN:AAF94173.1; GSPDB:GN001  
 A:Experimental source: serogroup O1, strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC1012  
 A:Map position: 1  
 C:Superfamily: conserved hypothetical protein H11688

Query Match  
 Best Local Similarity 2.1%; Score 9; DB 2; Length 230;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LVALPPGA 20  
 DB 184 LVALPPGA 192  
 |||||

RESULT 5  
 epidermal growth factor receptor - chicken  
 C:Species: Gallus gallus (chicken)  
 C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 18-Jun-1999  
 C:Accession: A42032  
 R:Flickinger, T.W.; Mahle, N.J.; Kung, H.J.  
 Mol. Cell. Biol. 12, 883-893, 1992  
 A>Title: An alternatively processed mRNA from the avian c-erbB gene encodes a soluble, t  
 A:Reference number: A42032; MUID:92123214; PMID:1732751  
 A:Accession: A42032  
 A:Molecule type: mRNA  
 A>Status: preliminary  
 A:Residues: 1-527 <FL>  
 A:Cross-references: GB:M77637; NID:g211737; PIDN:AAA48759.1; PID:g211738  
 A:Experimental source: liver  
 A>Note: sequence extracted from NCBI backbone (NCBI:76692, NCBI:76693)  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor

Query Match  
 Best Local Similarity 2.1%; Score 9; DB 2; Length 527;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 QCAAGCTGP 247  
 DB 245 QCAAGCTGP 253  
 |||||

RESULT 6  
 epidermal growth factor receptor - rat  
 A:Species: Rattus norvegicus (Norway rat)  
 C>Date: 25-Jan-1991 #sequence\_revision 25-Jan-1991 #text\_change 10-Oct-1997  
 C:Accession: A36325  
 R:Petich, L.A.; Harris, J.; Raymond, V.W.; Blasband, A.; Lee, D.C.; Earp, H.S.  
 Mol. Cell. Biol. 10, 2973-2982, 1990  
 A>Title: A truncated, secreted form of the epidermal growth factor receptor is encoded b  
 A:Reference number: A36325; MUID:90258888; PMID:2342466  
 A:Accession: A36325  
 A:Molecule type: mRNA  
 A>Status: preliminary  
 A:Residues: 1-644 <PE>  
 A:Cross-references: GB:M37394  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: alternative splicing; ATP; growth factor receptor

Query Match  
 Best Local Similarity 2.1%; Score 9; DB 2; Length 644;  
 Matches 9; Conservative 100.0%; Pred. No. 2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 QCAAGCTGP 247  
 DB 235 QCAAGCTGP 243  
 |||||

RESULT 7  
 epidermal growth factor receptor precursor - human  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
 C:Species: Homo sapiens (man)  
 C>Date: 15-Nov-1984 #sequence\_revision 27-Nov-1985 #text\_change 11-Jun-1999  
 C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143;  
 R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.  
 rg, P.H.  
 Nature 309, 418-425, 1984  
 A>Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression  
 A:Reference number: A00641; MUID:84219729; PMID:6328312  
 A:Accession: A00641  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <U>  
 A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924  
 A>Note: the authors translated the codon AAG for residue 540 as Asn  
 R:Shih, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985  
 A>Title: Characterization and sequence of the promoter region of the human epidermal  
 A:Reference number: A25772; MUID:85270438; PMID:2991899  
 A:Accession: A25772  
 A>Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-29 <ISH>  
 A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272  
 R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ulrich, A.; Waterfield, M.  
 Oncogene Res. 1, 375-396, 1987  
 A>Title: The human EGF receptor gene: structure of the 110 kb locus and identification  
 A:Reference number: S30024; MUID:88217333; PMID:3329716  
 A:Accession: S30024  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HA>  
 A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119  
 R:Haley, J.D.; Waterfield, M.D.  
 J. Biol. Chem. 266, 1746-1753, 1991  
 A>Title: Contributory effects of de Novo transcription and premature transcript termi  
 A:Reference number: A38672; MUID:91107677; PMID:1988448  
 A:Accession: A38672  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HAL>  
 A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271  
 R:Xu, Y.; Shih, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.;  
 Nature 309, 806-810, 1984  
 A>Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN  
 A:Reference number: A00642; MUID:84245835; PMID:6330563  
 A:Accession: A00642  
 A:Molecule type: mRNA  
 A:Residues: 1-150-187, 'KSVIOAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-  
 '798-799, 'ID', 802-811, 'R', 813-942 <XU>  
 A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF re  
 R:Lin, C.R.; Chen, W.S.; Krutger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma,  
 Science 224, 843-848, 1984  
 A>Title: Expression cloning of human EGF receptor complementary DNA: gene amplificati  
 A:Reference number: A43615; MUID:84196372; PMID:6326261  
 A:Accession: A43615  
 A:Molecule type: mRNA  
 A:Residues: 713-964 <LIN>  
 A:Experimental source: epidermoid carcinoma cell line A431  
 R:Slimmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.  
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984  
 A:Reference number: A23062; MUID:85046483; PMID:6093780  
 A:Accession: A23062  
 A:Molecule type: mRNA  
 A:Residues: 1028-1210 <SIM>

R:Weber, W.; Gill, G.N.; Speits, J.  
Science 224, 294-297, 1984  
A:Reference number: A05281; MUID:84172183; PMID:6324343  
A:Accession: A05281  
A:Molecule type: protein  
A:Residues: 25-30,'S',32-51,454-467 <WEB>  
R:Russo, M.W.; Lukasz, T.J.; Cohen, S.; Staros, J.V.  
J. Biol. Chem. 260, 5205-5208, 1985  
A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor  
A:Reference number: A60143; MUID:85182650; PMID:2965580  
A:Accession: A60143  
A:Molecule type: protein  
A:Residues: 740-744,'X',746-747 <RUS>  
R:Mroczkowski, B.; Mosig, G.; Cohen, S.  
Nature 309, 270-273, 1984  
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase  
A:Reference number: A38023; MUID:84191554; PMID:6325948  
A:Contents: annotation; receptor activity  
A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
R:Chen, W.S.; Lazzer, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.; Cell 59, 33-43, 1989  
A:Title: Functional independence of the epidermal growth factor receptor from a domain I  
A:Reference number: A33331; MUID:90093233; PMID:2780960  
A:Contents: annotation; internalization signal  
C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex  
C:Genetics:  
A:Gene: GDB:EGFR  
A:Cross-references: GDB:120610; OMIM:131550  
A:Map position: 7p12.3-7p12.1  
A:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase; signal transduction; tyrosine phosphorylation  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-1210/Product: EGF receptor #status predicted <MAT>  
F:25-645/Domain: extracellular #status predicted <EXT>  
F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>  
F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>  
F:646-668/Domain: transmembrane #status predicted <TM>  
F:668-1210/Domain: intracellular #status predicted <INT>  
F:710-975/Domain: protein kinase homology <KIN>  
F:718-726/Region: protein kinase ATP-binding motif  
F:999-1046/Region: coated-pit mediated internalization signal  
F:1047-1210/Region: inhibitory  
F:128,175,352,413,444,528,603/Binding site: carboxylate (Asn) (covalent) #status predicted <B>  
F:745/Active site: Lys #status experimental

Query Match	2.18	Score 9	DB 1	Length 1210
Best Local Similarity	100.0%	Pred. NO. 3.5		
Matches 9	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Oy 239 OCAAGCTGP 247  
Db 235 OCAAGCTGP 243

RESULT 8  
A53183  
epidermal growth factor receptor precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 18-Jun-1999  
C:Accession: A53183; A43818; S24942; A28941; S45325; I49643  
R:Luetkeke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.; Genes Dev. 8, 399-413, 1994  
A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor tyrosine kinase  
A:Reference number: A53183; MUID:94170986; PMID:8125255  
A:Accession: A53183  
A:Molecule type: mRNA  
A:Residues: 1-1210 <LUE>  
A:Cross-references: GB:U03425  
R:Avivi, I.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.  
Oncogene 6, 673-676, 1991  
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site  
A:Reference number: A43818; MUID:91232866; PMID:2030916  
A:Accession: A43818

```

A:Molecule type: mRNA
A:Residues: 1-714 <AVI>
A:Cross-references: GB:X59698
R:Eisinger, D.P.; Serreto, G.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942
A:Molecule type: mRNA
A:Residues: 969-971,'K',973-1115,'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylate
A:Reference number: A28941, MUID:88330814, PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694,'X',696-704,'L',706-707;989-992,'XX',995-996,'X',998-1000;1002-10
R:Hiibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Recept
A:Reference number: S45325
A:Accession: S45325
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-971,'K',973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:9488830; PIDN:CAA55587.1; PID:9488831
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mous
A:Reference number: 149643; MUID:93126380; PMID:7678348
A:Accession: 149643
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20,22-132 <RES>
A:Cross-references: GB:I06864; NID:g193001; PIDN:AAA53029.1; PID:g67201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor: protein kinase homology
C:Keywords: ATP: growth factor receptor; kinase-related transforming protein; phospho
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase ATP-binding motif
F:680,685/Binding site: phosphate (Thr) (covalent) #status experimental
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:969/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimenta
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 2.1%; Score 9; DB 2; Length 1210;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 QCAAGCTGP 247
|||
Db 235 QCAAGCTGP 243

RESULT 9
TVCHTV
epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Venn
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in
A:Reference number: A27720; MUID:88261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386

```

R.Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Ciltenden, L.B.; Raines, M.  
Cell 41, 719-726, 1985  
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro  
A:Reference number: A00643; MUID:85228222; PMID:2988784  
A:Accession: A00643  
A:Molecule type: mRNA  
A:Residues: 585-1223 <NIL>  
A:Cross-references: GB:M10066  
C:Genetics:  
A:Gene: erbB  
C:superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor  
specific protein kinase  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>  
F:31-1223/Domain: extracellular #status predicted <EXT>  
F:31-654/Domain: EGF receptor extracellular domain repeat <E1>  
F:397-610/Domain: EGF receptor extracellular domain repeat <E2>  
F:655-677/Domain: transmembrane #status predicted <TM>  
F:678-1223/Domain: intracellular #status predicted <INT>  
F:719-984/Domain: protein kinase homology <KIN>  
F:727-735/Region: protein kinase ATP-binding motif  
F:136,202,280,361,370,422,575,580,615,635/Binding site: carboxylate (Thr) (covalent) #  
F:192,650/Binding site: carboxylate (Ser) (covalent) #status predicted  
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
F:754/Active site: Lys #status predicted  
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 2.1%; Score 9; DB 1; Length 1223;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 239 QCAAGCTGP 247  
DB 242 QCAAGCTGP 250

RESULT 10  
JC5284  
carboxyl reductase (NADPH2) (EC 1.1.1.184), inducible - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 16-Apr-1997 #sequence\_revision 18-Jul-1997 #text\_change 03-Jun-2002  
C:Accession: S68982; PC2234; JC5284; PC2159; S52349  
R:Wernuth, B.; Maeder-Helmann, G.; Ernst, E.  
Eur. J. Biochem. 228, 473-479, 1995  
A:Title: Cloning and expression of carboxyl reductase from rat testis.  
A:Reference number: S68982; MUID:95220378; PMID:7705364  
A:Accession: S68982  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-277 <WER>  
A:Cross-references: EMBL:X84349; NID:g666086; PIDN:CA59088.1; PID:g666087  
R:Toft, E.; Soederstrom, M.; Ahlberg, M.B.; DePierre, J.W.  
Biochem. Biophys. Res. Commun. 201, 149-154, 1994  
A:Title: A novel 34kDa glutathione-binding protein in mature rat ovary.  
A:Reference number: PC2159; MUID:94256971; PMID:8198567  
A:Accession: PC2234  
A:Molecule type: protein  
A:Residues: 104-121,'X','123-134','D','136-137','H' <TOF>  
A:Experimental source: ovary  
R:Aoki, H.; Okada, T.; Mizutani, T.; Numata, Y.; Minegishi, T.; Miyamoto, K.  
Biochem. Biophys. Res. Commun. 230, 518-523, 1997  
A:Title: Identification of two closely related genes, inducible and noninducible carbony  
A:Reference number: JC5284; MUID:97167735; PMID:9015353  
A:Contents: ovary  
A:Accession: JC5284  
A:Molecule type: mRNA  
A:Residues: 1-140,'GM','143','R','145-235','T','237-238','E','240-277' <AOK>  
A:Cross-references: DBJ:B89069; NID:g1906811; PIDN:BA19007.1; PID:g1906812  
C:Comment: This enzyme is a cytosolic, monomeric oxidoreductase that catalyzes the NADPH  
C:Genetics:  
A:Gene: ICR  
C:superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: cytosol; monomer; NADP; oxidoreductase  
F:6-185/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 1.9%; Score 8; DB 2; Length 277;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 LALLPGA 20  
DB 254 LALLPGA 261

RESULT 11  
B83298  
conserved hypothetical protein PA2771 [imported] - Pseudomonas aeruginosa (strain PAO  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83298  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: B83298  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-341 <STO>  
A:Cross-references: GB:AE004705; GB:AE004091; NID:g9948851; PIDN:AAG06159.1; GSPDB:GN  
A:Experimental source: strain PAO1  
C:Genetics:  
A:Gene: PA2771

Query Match 1.9%; Score 8; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 140 LQRLSLE 147  
DB 158 LQRLSLE 165

RESULT 12  
D95067  
phenylalanyl-tRNA synthetase, alpha chain [imported] - Streptococcus pneumoniae (stra  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C:Accession: D95067  
R:Petelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Redune, D.; Holtzapp  
son, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: D95067  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-348 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK74733.1; PID:g14972054; GSPDB:GN00164; TIGR  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0579  
C:superfamily: phenylalanine-tRNA ligase alpha chain

Query Match 1.9%; Score 8; DB 2; Length 348;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 144 SLTEILKG 151  
DB 43 SLTEILKG 50

## RESULT 13

C97935  
phenylalanine-tRNA ligase (EC 6.1.1.20) [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence-revision 22-oct-2001 #text\_change 03-Jun-2002  
C:Accession: C97935  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E  
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: C97935  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-375 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99311.1; PID:g15458079; GSPDB:GN00174  
C:Genetics:  
A:Gene: phs  
C:Superfamily: phenylalanine-tRNA ligase alpha chain  
C:Keywords: ligase

Query Match 1.9%; Score 8; DB 2; Length 375;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 SLTEILKG 151  
Db 70 SLTEILKG 77

## RESULT 14

H70922  
hypothetical protein RV3130c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence-revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: H70922  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: H70922  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-463 <COL>  
A:Cross-references: GB:Z95150; GB:AL123456; NID:g3250708; PIDN:CAR08399.1; PID:el301034;  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3130c

Query Match 1.9%; Score 8; DB 2; Length 463;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 PLQRLRIV 102  
Db 315 PLQRLRIV 322

## RESULT 15

C70311  
hypothetical protein ag\_116 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence-revision 08-May-1998 #text\_change 09-Jun-2000  
C:Accession: C70311  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O  
V.  
Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: C70311

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-520 <AQF>

A:Cross-references: GB:AE000674; NID:g2982850; PIDN:AAC06482.1; PID:g2982861; GB:AE00

A:Experimental source: strain VF5

C:Genetics:

A:Gene: ag\_116

C:Superfamily: Aquifex aeolicus hypothetical protein ag\_116

Query Match 1.9%; Score 8; DB 2; Length 520;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 387 PLSPTSVP 394  
Db 2 PLSPTSVP 9

Search completed: March 4, 2003, 12:47:43  
Job time : 49.1165 secs.



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:37:16 ; Search time 28.6064 Seconds  
(without alignments)  
607.506 Million cell updates/sec

Title: US-09-234-208B-2

Perfect score: 419  
Sequence: 1 MELALCRMGDLALLPPGA.....VGRGPPDAHVAVNLRYEG 419

Scoring table: OLIGO  
Gapop 60.0 , Gapept 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	340	81.1	1255	1	ERB2_HUMAN
2	67	16.0	1254	1	ERB2_MESAU
3	52	12.4	1257	1	ERB2_RAT
4	9	2.1	230	1	RNFE_VIBCH
5	9	2.1	703	1	EGFR_CHICK
6	9	2.1	1210	1	EGFR_HUMAN
7	9	2.1	1210	1	EGFR_MOUSE
8	8	1.9	276	1	DHCA_RAT
9	8	1.9	463	1	YV30_MYCTU
10	8	1.9	725	1	NCA2_MOUSE
11	8	1.9	858	1	NCA1_RAT
12	8	1.9	1115	1	SECE_SULSO
13	7	1.7	62	1	SECE_MOUSE
14	7	1.7	104	1	Y164_ARCFU
15	7	1.7	111	1	NMDE_HUMAN
16	7	1.7	114	1	GON2_TUPGB
17	7	1.7	173	1	NM6M_CHICK
18	7	1.7	189	1	YAT5_RHOB
19	7	1.7	192	1	UREE_ALCEU
20	7	1.7	207	1	YPOP_BACSU
21	7	1.7	220	1	GF33_CHICK
22	7	1.7	221	1	SDFL_HUMAN
23	7	1.7	224	1	TCFD_SALTY
24	7	1.7	228	1	YU02_HUMAN
25	7	1.7	248	1	DSBG_ECO57
26	7	1.7	248	1	DSBG_ECOLI
27	7	1.7	269	1	NMFL_AZOCH
28	7	1.7	271	1	HMGL_TRYBR
29	7	1.7	287	1	YCKE_BACSU
30	7	1.7	306	1	SRP1_SYNP7
31	7	1.7	307	1	MRP1_MYCLE
32	7	1.7	313	1	ENVI_MOUSE
33	7	1.7	334	1	G3P1_BACSU

34	7	1.7	341	1	XERD_SELRU
35	7	1.7	357	1	MLTC_HAEIN
36	7	1.7	365	1	Y271_YEAST
37	7	1.7	379	1	IRKA_RAT
38	7	1.7	380	1	YMH7_CABEL
39	7	1.7	383	1	PODO_HUMAN
40	7	1.7	396	1	YD18_YEAST
41	7	1.7	399	1	BRS3_SHEEP
42	7	1.7	400	1	RENI_SHEEP
43	7	1.7	423	1	ANCT_HUMAN
44	7	1.7	443	1	CGUH_XANMA
45	7	1.7	443	1	SOX3_HUMAN

## ALIGNMENTS

RESULT 1  
ID ERB2\_HUMAN STANDARD; PRT: 1255 AA.  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell  
DE surface receptor HER2) (MLN 19).  
GN ERB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-86118663; PubMed-3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.;  
RT "Similarity of protein encoded by the human c-erbB-2 gene to  
RT epidermal growth factor receptor.";  
RL Nature 319:230-234(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-86016729; PubMed-2999974;  
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Francke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139(1985).  
RN [3]  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE-86016729; PubMed-2999974;  
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the  
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a  
RL human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
RN [4]  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE-93194196; PubMed-8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429(1993).  
CC -i- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
CC ALTHOUGH IT DOES NOT INTERACT WITH IT ALONE. GP30 IS A  
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
CC ALPHA AND AMPHIREGULIN.  
CC -i- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -i- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC -i- SUBCELLULAR LOCATION: Type I membrane protein.

CC -i- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
CC RESIDUES (BY SIMILARITY).  
CC -i- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
CC -i- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M11767; AAA35808.1; -  
CC EMBL: M11761; AAA35808.1; JOINED.  
CC EMBL: M11762; AAA35808.1; JOINED.  
CC EMBL: M11763; AAA35808.1; JOINED.  
CC EMBL: M11764; AAA35808.1; JOINED.  
CC EMBL: M11765; AAA35808.1; JOINED.  
CC EMBL: M11766; AAA35808.1; JOINED.  
CC EMBL: M11730; AAA35808.1; -  
CC EMBL: M12036; AAA35978.1; -  
CC EMBL: X03363; CAA27060.1; -  
CC PIR: A25491; A25491.  
CC PIR: A24571; A24571.  
CC HSP: P11362; 1FGK.  
CC Genew: HGNC:3430; ERBB2.  
CC MIM: 164870; -  
CC InterPro: IPR000494; EGFR\_L\_domain.  
CC InterPro: IPR000719; Euk\_pkinase.  
CC InterPro: IPR002174; Furin-like.  
CC InterPro: IPR001245; Tyr\_pkinase.  
CC InterPro: IPR004019; YLP\_motif.  
CC Pfam: PF00069; pkinase; 1.  
CC Pfam: PF00757; Furin-like; 1.  
CC Pfam: PF01030; Recep\_L\_domain; 2.  
CC Pfam: PF02757; YLP; 2.  
CC Prodom: PD000001; Euk\_pkinase; 1.  
CC SMART: SM00261; FU; 3.  
CC SMART: SM00219; TYRC; 1.  
CC PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
CC PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
CC PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
CC Transmembrane: Glycoprotein: Multigene family: Receptor: Signal:  
CC Transferrase: Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
CC Polymorphism.  
CC SIGNAL: 1 21  
CC CHAIN: 22 1255 POTENTIAL.  
CC DOMAIN: 22 652 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
CC TRANSMEM: 653 675 EXTRACELLULAR (POTENTIAL).  
CC DOMAIN: 676 1255 POTENTIAL.  
CC DOMAIN: 720 987 CYTOPLASMIC (POTENTIAL).  
CC NP\_BIND: 726 987 PROTEIN KINASE.  
CC BINDING: 753 753 ATP (BY SIMILARITY).  
CC ACT\_SITE: 845 845 ATP (BY SIMILARITY).  
CC DISULFID: 195 204 BY SIMILARITY.  
CC DISULFID: 199 212 BY SIMILARITY.  
CC DISULFID: 220 227 BY SIMILARITY.  
CC DISULFID: 224 235 BY SIMILARITY.  
CC DISULFID: 236 244 BY SIMILARITY.  
CC DISULFID: 240 252 BY SIMILARITY.  
CC DISULFID: 255 264 BY SIMILARITY.  
CC DISULFID: 268 295 BY SIMILARITY.  
CC DISULFID: 299 311 BY SIMILARITY.  
CC DISULFID: 315 331 BY SIMILARITY.  
CC DISULFID: 334 338 BY SIMILARITY.  
CC DISULFID: 511 520 BY SIMILARITY.  
CC DISULFID: 515 528 BY SIMILARITY.  
CC DISULFID: 531 540 BY SIMILARITY.  
CC DISULFID: 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
FT DISULFID 567 584 BY SIMILARITY.  
FT DISULFID 587 596 BY SIMILARITY.  
FT DISULFID 600 623 BY SIMILARITY.  
FT DISULFID 626 634 BY SIMILARITY.  
FT DISULFID 630 642 BY SIMILARITY.  
FT MOD\_RES 1139 1139 BY SIMILARITY.  
FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CAROAMD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CAROAMD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CAROAMD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CAROAMD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CAROAMD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CAROAMD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CAROAMD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 654 654 I -> V.  
FT VARIANT 654 654 /FTID=VAR\_004077.  
FT VARIANT 655 655 I -> V.  
FT CONFLICT 1170 1170 /FTID=VAR\_004078.  
FT SEQUENCE 1255 AA; 137909 MW; 39E9DFDA04DCF962 CRC64;  
SQ  
Query Match 81.1%; Score 340; DB 1; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MELALCRWGLLALLPPAASVCTGDMKRLPASPETHLMDLRHLVGGCQVQGNL 60  
DB 1 MELALCRWGLLALLPPAASVCTGDMKRLPASPETHLMDLRHLVGGCQVQGNL 60  
QY 61 ELTYLEPTNASLSFLDDIOEVGCVLIAHNOVRVPLQRLIRIVGSQLFEDNYALAVLNDG 120  
DB 61 ELTYLEPTNASLSFLDDIOEVGCVLIAHNOVRVPLQRLIRIVGSQLFEDNYALAVLNDG 120  
QY 121 DPLNNTPTVGTASPGGLRDLRLSTLEIKKGVLLQNRNQLCYDTILMKDLEHNNQLA 180  
DB 121 DPLNNTPTVGTASPGGLRDLRLSTLEIKKGVLLQNRNQLCYDTILMKDLEHNNQLA 180  
QY 121 DPLNNTPTVGTASPGGLRDLRLSTLEIKKGVLLQNRNQLCYDTILMKDLEHNNQLA 180  
DB 121 DPLNNTPTVGTASPGGLRDLRLSTLEIKKGVLLQNRNQLCYDTILMKDLEHNNQLA 180  
QY 181 LTLIDTNRSRACHPCSPMKSGSRWGESSESDCSLTRYCAGGACRKGPLPTDCCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPMKSGSRWGESSESDCSLTRYCAGGACRKGPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLEHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLEHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300  
QY 301 YNVLSTDVSGSCLVCPRLHNOVTAEDGTQRCERCKSPCAR 340  
DB 301 YNVLSTDVSGSCLVCPRLHNOVTAEDGTQRCERCKSPCAR 340  
RESULT 2  
ERR2\_MESAU STANDARD; PRT; 1254 AA.  
ID ERR2\_MESAU  
AC Q60553;  
DE 15-DEC-1998 (Rel. 37, Created)  
DE 15-DEC-1998 (Rel. 37, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbb-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).  
DE ERBB2 OR NEU.  
OS Mesocricetus auratus (Golden hamster).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Mesocricetus.  
OC NCBI\_TaxId=10036;  
OX NCBI\_TaxId=10036;  
RN 11  
RP SEQUENCE FROM N.A.  
RC TISSUE=Nerve;  
RX MEDLINE=94193007; PubMed=7908275;  
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,  
RA Yamazaki Y., Ishikawa T.;  
RA "Cloning and activation of the Syrian hamster neu proto-oncogene";

RL Gene 140:251-255(1994).  
 CC -i- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).  
 CC -i- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -i- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -i- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -i- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -i- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D16295; BAA03801.1; -  
 DR HSSP: P11362; 1FGR.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00261; EU\_3.  
 DR SMART: SM00219; tyrc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Proto-oncogene; Disease mutation.  
 KW SIGNAL  
 FT CHAIN 1 21  
 FT 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT TRANSMEM 653 675 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 676 1254 POTENTIAL.  
 FT 158 368 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 472 644 CYS-RICH.  
 FT 720 987 PROTEIN KINASE.  
 FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
 FT BINDING 753 753 ATP (BY SIMILARITY).  
 FT ACT\_SITE 845 845 BY SIMILARITY.  
 FT 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 268 295 BY SIMILARITY.  
 FT 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.  
 FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT MOD\_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).  
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).  
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;  
 Query Match 16.0%; Score 67; DB 1; Length 1254;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-59;  
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 231 LPDCCHEQCAAGCTGPKHSDCIACLFHNSGICELHCPALVYNTDFESMNPGRYT 290  
 DB 231 LPDCCHEQCAAGCTGPKHSDCIACLFHNSGICELHCPALVYNTDFESMNPGRYT 290  
 QY 291 FGASCVT 297  
 DB 291 FGASCVT 297  
 RESULT 3  
 ID ERB2\_RAT STANDARD; PRT; 1257 AA.  
 AC P06494;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbb-2 precursor (EC 2.7.1.112)  
 DE (p18etrb2) (NEU proto-oncogene) (C-erbb-2) (Epidermal growth factor  
 DE receptor-related protein).  
 GN ERB2 OR NEU.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neuroblastoma;  
 RX MEDLINE=86118662; PubMed=3945311;  
 RA Bargman C.I., Hung M.-C., Weinberg R.A.;  
 RT "The neu oncogene encodes an epidermal growth factor receptor-related  
 RT protein.";  
 RL Nature 319:226-230(1986).  
 RN [2]  
 RP SEQUENCE OF 852-905 FROM N.A.  
 RC TISSUE=sciatic nerve;  
 RX MEDLINE=91222560; PubMed=2025425;  
 RA Lai C., Lemke G.;  
 RT "An extended family of protein-tyrosine kinase genes differentially  
 RT expressed in the vertebrate nervous system.";  
 RL Neuron 6:691-704(1991).  
 RN [3]  
 RP STRUCTURE BY NMR OF 650-668.  
 RX MEDLINE=9215181; PubMed=1346763;  
 RA Gullick W.J., Bottomley A.C., Lofers F.J., Doak D.G., Mulvey D.,  
 RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;  
 RT "Three dimensional structure of the transmembrane region of the proto-  
 RT oncogenic and oncogenic forms of the neu protein.";  
 RL EMBO J. 11:43-48(1992).  
 CC -i- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN.  
 CC -i- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -i- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.  
 CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.  
 CC -i- SUBCELLULAR LOCATION: Type I membrane protein.

```

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X03362; CA27059.1; ALT_INIT.
DR PIR: A24562; TVRTND.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk.pkinase.
DR InterPro: IPR002174; Furlin-like.
DR InterPro: IPR001245; Tyr.pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00757; Furlin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; F0; 3.
DR SMART: SM00219; TYRC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
DR Transmembrane: Tyrosine-protein kinase; ATP-binding; Phosphorylation;
DR Proto-oncogene; Disease mutation.
KW SIGNAL.
FT CHAIN 1 21
FT DOMAIN 22 1257
FT TRASMEM 655 674
FT DOMAIN 678 1257
FT DOMAIN 159 369
FT DOMAIN 473 646
FT DOMAIN 722 989
FT NP_BIND 728 736
FT BINDING 755 755
FT ACT_SITE 847 847
FT DISULFID 196 205
FT DISULFID 200 213
FT DISULFID 221 228
FT DISULFID 225 236
FT DISULFID 237 245
FT DISULFID 241 253
FT DISULFID 256 265
FT DISULFID 269 296
FT DISULFID 300 312
FT DISULFID 316 332
FT DISULFID 335 339
FT DISULFID 513 522
FT DISULFID 517 530
FT DISULFID 533 542
FT DISULFID 546 562
FT DISULFID 565 578
FT DISULFID 569 586
FT DISULFID 589 598
FT DISULFID 602 625
FT DISULFID 628 636
FT DISULFID 632 644
FT MOD_RES 1141 1141
FT MOD_RES 1250 1250
FT CARBOHYD 68 68
FT CARBOHYD 188 188
FT CARBOHYD 260 260
FT CARBOHYD 532 532
FT CARBOHYD 573 573
FT CARBOHYD 631 631

```

```

FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;
Query Match
Best Local Similarity 12.4%; Score 52; DB 1; Length 1257;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 231 LPTDCHEQCAAGCTGPKHSDCLAFPHNSGICELHCPALVYNTDFESM 282
DB 232 LPTDCHEQCAAGCTGPKHSDCLAFPHNSGICELHCPALVYNTDFESM 283
RESULT 4
RNFE_VIBCH
ID RNFE_VIBCH STANDARD; PRT; 230 AA.
AC 09KT91;
DT 15-JUN-2002 (rel. 41, Last sequence update)
DE 15-JUN-2002 (rel. 41, Last annotation update)
DE Electron transport complex protein rnfe.
GN RNFE OR VC1012.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- SUBUNIT: Composed of at least six subunits; rnfa, rnfb, rnfc,
CC rnfd, rnfe and rnfg (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE NORDE/RNFAE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE004183; AAF94173.1; -.
DR TIGR: VC1012; -.
DR InterPro: IPR003677; Rnf_Nqr.
DR Pfam: PF02508; Rnf-Nqr; 1.
KW Electron transport; Transmembrane; Inner membrane; Complete proteome.
FT TRASMEM 39 59
FT TRASMEM 69 89
FT TRASMEM 93 113
FT TRASMEM 124 144
FT TRASMEM 182 202
SQ SEQUENCE 230 AA; 24710 MW; 262D4D792044D769 CRC64;
Query Match
Best Local Similarity 2.1%; Score 9; DB 1; Length 230;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 LVALPPGA 20
DB 184 LVALPPGA 192

```

```

RESULT 5
EGFR_CHICK
ID EGFR_CHICK STANDARD: PRT: 703 AA.
AC P1387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: M20386; AAA8760.1;
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU; 4.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; PARTIAL.
DR Transmembrane: Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1
FT CHAIN 1
FT DOMAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT TRANSMEM 655 667 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 668 >703 POTENTIAL.
FT DISULFID 197 206 CYTOPLASMIC (POTENTIAL).
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.

```

```

FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 703
SQ SEQUENCE 703 AA; 77427 MW; AFD2DE1B735A690 CRC64;

Query Match      2.1%; Score 9; DB 1; Length 703;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 QCAAGCTGP 247
   |||||
Db 242 QCAAGCTGP 250

RESULT 6
EGFR_HUMAN
ID EGFR_HUMAN STANDARD: PRT: 1210 AA.
AC P00533; P06268; 014225; Q9JUMD7; Q9JUM8; Q9JUM5; Q92795; 000732;
AC 000688; Q9B2S2; Q9H2C9; Q9GZX1; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE protein-tyrosine kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells."
RT Nature 309:418-425(1984).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE-Placenta;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta."
RL Mol. Reprod. Dev. 41:149-156(1995).
RN (3)
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE-Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Mahle N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth

```

RT factor receptor gene encodes a truncated form of the receptor.";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE-Placenta;  
 RX MEDLINE=97256547; PubMed=9103388;  
 RT Ilekis J.V., Garfili J., Niederberger C., Scoccia B.;  
 "Expression of a truncated epidermal growth factor receptor-like  
 protein (TEGFR) in ovarian cancer.";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RC TISSUE-Placenta;  
 RX MEDLINE=21100872; PubMed=11161793;  
 RT Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
 Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.,  
 Mailhe N.J.;  
 "Comparative genomic sequence analysis and isolation of human and  
 mouse alternative EGF transcripts encoding truncated receptor  
 isoforms.";  
 RL Genomics 71:1-20(2001).  
 RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schehl C.M.,  
 Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,  
 Mailhe N.J.;  
 "Human and mouse alternative EGF transcripts encoding only the  
 extracellular domain of the receptor";  
 RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krutger W., Stolarsky L.S., Weber W.,  
 Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 "Expression cloning of human EGF receptor complementary DNA: gene  
 amplification and three related messenger RNA products in A431  
 cells.";  
 RL Science 224:843-848(1984).  
 RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 Roe B.A., Merlino G.T., Pastan I.;  
 "Human epidermal growth factor receptor cDNA is homologous to a  
 variety of RNAs overproduced in A431 carcinoma cells.";  
 RL Nature 309:806-810(1984).  
 RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,  
 O'Malley B.W.;  
 "Isolation of an evolutionarily conserved epidermal growth factor  
 receptor cDNA from human A431 carcinoma cells.";  
 RL Biochem. Biophys. Res. Commun. 124:123-132(1984).  
 RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 Waterfield M.D.;  
 "The human EGF receptor gene: structure of the 110 kb locus and  
 identification of sequences regulating its transcription.";  
 RL Oncogene Res. 1:375-396(1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 "Contributory effects of de novo transcription and premature  
 transcript termination in the regulation of human epidermal growth  
 factor receptor proto-oncogene RNA synthesis.";  
 RL J. Biol. Chem. 266:1746-1753(1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., the promoter region of the human  
 epidermal growth factor receptor gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN [14]  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor  
 and supercoiled DNA.";  
 RL Nature 309:270-273(1984).  
 RN [15]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,  
 Howk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 receptor and HHR23neu are located in their carboxyl-terminal tails.  
 Identification of a novel site in EGF receptor.";  
 RL J. Biol. Chem. 264:10667-10671(1989).  
 RN [16]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96398132; PubMed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 the epidermal growth factor receptor expressed in Chinese hamster  
 ovary fibroblasts.";  
 RL Growth Factors 13:121-132(1996).  
 RN [17]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731688;  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 Asn-X-Cys sequence of recombinant human epidermal growth factor  
 receptor.";  
 RL J. Biochem. 127:65-72(2000).  
 RN [18]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; PubMed=9556602;  
 RA Abe Y., Otake M., Inagaki F., Lax I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor.";  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RN [19]  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 mitogens.";  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 RN [20]  
 RP FUNCTION: Receptor for EGF, but also for other members of the EGF  
 family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 EGF-like growth factor, GP30 and vaccinia virus growth factor. It  
 is involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA



```

FT DISULFID 295 307 BY SIMILARITY.
FT DISULFID 311 326 BY SIMILARITY.
FT DISULFID 329 333 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 582 579 BY SIMILARITY.
FT DISULFID 595 591 BY SIMILARITY.
FT DISULFID 620 617 BY SIMILARITY.
FT DISULFID 624 628 BY SIMILARITY.
FT DISULFID 636 636 BY SIMILARITY.
FT MOD.RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD.RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD.RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD.RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD.RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DFD2F5 CRC64;

Query Match 2.1%; Score 9; DB 1; Length 1210;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 239 QCAAGCTGP 247
Db 235 QCAAGCTGP 243

RESULT 8
DHCA_RAT STANDARD: PRT: 276 AA.
AC P47727;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbonyl reductase [NADPH] 1 (EC 1.1.1.184) (NADPH-dependent carbonyl reductase 1).
GN CBRL OR CBR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RA MEDLINE=95220378; PubMed=705364;
RA Wermuth B., Meeder-Heinmann G., Ernst E.;
RT "Cloning and expression of carbonyl reductase from rat testis.";
RL Eur. J. Biochem. 228:473-479(1995).
CC -1- CATALYTIC ACTIVITY: R-CHOH-R' + NADP(+) = R-CO-R' + NADPH.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDS) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X83439; CAA59088.1; -.
CC EMBL; X95986; CAA65230.1; -.
CC HSSP; P50162; IAEI.
CC InterPro: IPR002198; ADH_short.
CC Pfam: PF00106; adh_short; 1.
CC PROSITE; P500061; ADH_SHORT; 1.
CC Oxidoreductase; NADP; Acetylation.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD.RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 9 33 NADP (BY SIMILARITY).
FT ACT_SITE 193 193 BY SIMILARITY.
SQ SEQUENCE 276 AA; 30447 MW; 99B9E77C5E2922AB CRC64;

Query Match 1.9%; Score 8; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 LALLPGGA 20
Db 253 LALLPGGA 260

RESULT 9
YV30_MYCTU STANDARD: PRT: 463 AA.
AC Y07035;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein RV3130c.
GN RV3130C OR MT3216 OR MTCY03A2.28 OR MTCY164.41c.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OX Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tebala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,
RA Sultson J.F., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE UPF0089 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: 295150; CAB08399.1; -.  
DR EMBL: AE007137; AAK47554.1; -.  
DR TIGR: MT3216; -.  
DR Tuberculin; Rv3130c; -.  
DR InterPro: IPR004255; UPF0089.  
DR Pfam: PF03007; UPF0089; 1.  
KW Hypothetical protein: Complete proteome.  
FT CONFLICT 268 268 L -> F (IN REF. 2).  
SQ SEQUENCE 463 AA; 50720 MW; 48CC60A8670F1D6B CRC64;  
  
Query Match 1.9%; Score 8; DB 1; Length 463;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 95 PLOLRATV 102  
DB 315 PLOLRATV 322  
-----  
RESULT 10  
NCB2\_MOUSE  
ID NCB2\_MOUSE STANDARD; PRT; 725 AA.  
AC P13594; Q61950;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Neutral cell adhesion molecule 1, 120 kDa isoform precursor (N-CAM 120)  
DE (NCAM-120).  
GN NCAM1 OR NCAM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCB1\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RX MEDLINE=87246524; PubMed=3595563;  
RA Barthele D., Santoni M.-J., Wille W., Ruppert C., Caix J.-C.,  
RT Hirsch M.-R., Fontecilla-Camps J.-C., Goridis C.;  
RT Isolation and nucleotide sequence of mouse NCAM cDNA that codes for  
RT a Mr 79,000 polypeptide without a membrane-spanning region.;  
RL EMBJ. 6:907-914(1987).  
RN [2]  
RP SEQUENCE OF 20-700 FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=89251563; PubMed=2721486;  
RA Santoni M.-J., Barthele D., Vopper G., Boned A., Goridis C., Wille M.;  
RT Differential exon usage involving an unusual splicing mechanism  
RT generates at least eight types of NCAM cDNA in mouse brain.;  
RL EMBJ. 8:385-392(1989).  
RN [3]  
RP SEQUENCE OF 642-725 FROM N.A.  
RC MEDLINE=88283628; PubMed=3396534;  
RX Barthele J.A., Chaix J.C., Steinmetz M., Goridis C.;  
RT Differential splicing and alternative polyadenylation generates  
RT distinct NCAM transcripts and proteins in the mouse.;  
RL EMBJ. 7:625-632(1988).  
RN [4]  
RP SEQUENCE OF 20-36.  
RX MEDLINE=86140120; PubMed=3512556;  
RA Rougon G., Marshak D.R.;  
RT Structural and immunological characterization of the amino-terminal  
RT domain of mammalian neural cell adhesion molecules.;  
RL J. Biol. Chem. 261:3396-3401(1986).  
CC -I- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF  
CC NEURITES, ETC.  
CC -I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -I- ALTERNATIVE PRODUCTS: 3 isoforms; N-CAM 180 (AC P13595), N-CAM 140

CC (AC P13595) and N-CAM 120 (shown here); are produced by  
CC alternative splicing.  
CC -I- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC -I- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: Y00051; CAA68263.1; -.  
DR EMBL: X15049; CAA33148.1; ALT\_SEQ.  
DR EMBL: X07195; CAA30173.1; -.  
DR PIR: A29673; IJMSNG.  
DR MGD: MGI:97281; Ncam1.  
DR InterPro: IPR003961; FN\_III.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003598; Ig\_C2.  
DR Pfam: PF00041; fn3; 2.  
DR Pfam: PF00047; Ig\_5.  
DR SMART: SM00060; FN3; 2.  
DR SMART: SM00408; IGC2; 5.  
KW Cell adhesion; Glycoprotein; Repeat; Alternative splicing;  
KW Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor.  
FT SIGNAL 1 19  
FT CHAIN 20 725  
FT FT  
FT DOMAIN 34 103  
FT DOMAIN 132 196  
FT DOMAIN 228 295  
FT DOMAIN 323 393  
FT DOMAIN 420 487  
FT DOMAIN 519 596  
FT DOMAIN 625 692  
FT DOMAIN 152 165  
FT DOMAIN 161 165  
FT DISULFID 41 96  
FT DISULFID 139 189  
FT DISULFID 235 288  
FT DISULFID 330 386  
FT DISULFID 427 480  
FT CARBOHYD 222 222  
FT CARBOHYD 316 316  
FT CARBOHYD 348 348  
FT CARBOHYD 424 424  
FT CARBOHYD 450 450  
FT CARBOHYD 479 479  
FT CONFLICT 261 268  
FT CONFLICT 273 273  
FT CONFLICT 354 355  
FT CONFLICT 549 549  
FT CONFLICT 572 572  
FT CONFLICT 575 575  
FT CONFLICT 589 594  
FT CONFLICT 600 602  
FT CONFLICT 657 657  
SQ SEQUENCE 725 AA; 80296 MW; C2AE8B8461C8B2F CRC64;  
  
Query Match 1.9%; Score 8; DB 1; Length 725;  
Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 322 VTAEDGTQ 329  
DB 98 VTAEDGTQ 105  
-----  
RESULT 11  
NCAL\_RAT





```

RESULT 14
ID Y164.ARCFU STANDARD: PRT: 104 AA.
AC 028415:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1864.
AF1864.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RA MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervatage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- SIMILARITY: Strong, to A.aeolicus AQ_377.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: AE000974; AAB89396.1; -
CC DR TIGR: AF1864; -
CC KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 104 AA; 11845 MW; DC80B36CF3F5C153 CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 104;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 EILKGV 153
DB 95 EILKGV 101

RESULT 15
MADE_HUMAN
ID MADE_HUMAN STANDARD: PRT: 111 AA.
AC 000984:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE p75NTR-associated cell death executor (Nerve growth factor receptor
DE associated protein 1) (Ovarian granulosa cell 13.0 kDa protein HGR74).
GN NGFRAP1 OR MADE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=91025550; PubMed=2171551;

```

```

RA Rapp G., Freudenstein J., Klaudiny J., Mucha J., Wempe F., Zimmer M.,
RA Scheit K.H.;
RT "Characterization of three abundant mRNAs from human ovarian
RT granulosa cells."
RT DNA Cell Biol. 9:479-485(1990).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC MEDLINE=20298829; PubMed=10764727;
RA Mukai J., Hachiya T., Shoji-Hoshino S., Kimura M.T., Nadano D.,
RA Suvanto P., Hanacka T., Li Y., Irie S., Greene L.A., Sato T.A.;
RT "MADE, a p75NTR-associated cell death executor, is involved in signal
RT transduction mediated by the common neurotrophin receptor p75NTR."
RT J. Biol. Chem. 275:17566-17570(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May be a signaling adaptor molecule involved in p75NTR-
CC mediated apoptosis induced by NGF. May play an important role in
CC the pathogenesis of neurodegenerative diseases.
CC -1- SUBUNIT: Binds to the death domain of p75NTR/NGFR.
CC -1- TISSUE SPECIFICITY: FOUND IN OVARIAN GRANULOSA CELLS, TESTIS,
CC PROSTATE AND SEMINAL VESICLE TISSUE.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: M38188; AAA63232.1; -
CC DR EMBL: AF187064; AAF75129.1; -
CC DR EMBL: BC003190; AAH03190.1; -
CC DR PIR: C35826; C35826.
CC DR Genew: HGNC:13388; NGFRAP1.
CC DR MIM: 300361; -
CC DR InterPro: IPR001230; Prenyl-site.
CC KW Apoptosis.
SQ SEQUENCE 111 AA; 12958 MW; 29AA0573282C933E CRC64;

Query Match
Best Local Similarity 1.7%; Score 7; DB 1; Length 111;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 LRELQLR 143
DB 81 LRELQLR 87

Search completed: March 4, 2003, 12:45:38
Job time : 30.6064 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:42:02 ; Search time 46.2751 Seconds  
(without alignments)  
1865.663 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 419  
Sequence: 1 MELALACRMGLLALLPPGA.....VGRGPDPAHVAVNLSEREG 419

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPREMBL\_21.\*  
1: sp-archaea:\*  
2: sp-bacteria:\*  
3: sp-fungi:\*  
4: sp-human:\*  
5: sp-invertebrate:\*  
6: sp-mammal:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*  
15: sp-virus:\*  
16: sp-bacteriap:\*  
17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	345	82.3	419	4	Q9UK79
2	67	16.0	149	6	Q9BG66
3	66	15.8	1259	6	Q9BG66
4	9	2.1	478	11	Q9ES60
5	9	2.1	527	13	Q9SH2
6	9	2.1	599	13	Q9SH2
7	9	2.1	643	11	Q9ERV6
8	9	2.1	655	11	Q9WV5
9	9	2.1	1209	11	Q9GX70
10	9	2.1	1210	11	Q9EP98
11	8	1.9	48	11	Q63365
12	8	1.9	63	11	Q9QZ16
13	8	1.9	63	11	Q9QZ16
14	8	1.9	217	5	Q9N6W6
15	8	1.9	277	11	Q9N6W6
16	8	1.9	277	11	Q924V3

17	8	1.9	280	13	Q8UVJ2	Q8UVJ2 oreochromis
18	8	1.9	341	16	Q91072	Q91072 pseudomonas
19	8	1.9	348	16	Q9TS36	Q9TS36 streptococcus
20	8	1.9	481	10	Q945M4	Q945M4 arabidopsis
21	8	1.9	520	16	Q66518	Q66518 aquilex aeo
22	8	1.9	583	16	Q99RE2	Q99RE2 staphylococcus
23	8	1.9	605	11	Q921P2	Q921P2 mus musculus
24	8	1.9	865	10	Q22774	Q22774 arabidopsis
25	8	1.9	879	10	Q41010	Q41010 pisum sativ
26	8	1.9	1436	17	Q96ZJ5	Q96ZJ5 sulfobolus
27	8	1.9	1469	10	Q91KRL	Q91KRL pisum sativ
28	8	1.9	1503	10	Q81283	Q81283 arabidopsis
29	8	1.9	1503	10	Q81283	Q81283 arabidopsis
30	7	1.7	27	4	Q9HD19	Q9HD19 homo sapien
31	7	1.7	39	5	Q8WUJ2	Q8WUJ2 plasmodium
32	7	1.7	61	10	Q8S9S5	Q8S9S5 oryza sativ
33	7	1.7	67	12	Q84264	Q84264 european el
34	7	1.7	114	11	Q9D1N5	Q9D1N5 mus musculus
35	7	1.7	117	2	Q8WMI3	Q8WMI3 pseudomonas
36	7	1.7	120	11	Q9JIT2	Q9JIT2 rattus norv
37	7	1.7	121	16	Q8YZW8	Q8YZW8 anabaena sp
38	7	1.7	124	11	Q9CWN9	Q9CWN9 mus musculus
39	7	1.7	124	11	Q9WT29	Q9WT29 mus musculus
40	7	1.7	146	15	Q8UJ38	Q8UJ38 human immun
41	7	1.7	146	15	Q8UJ38	Q8UJ38 human immun
42	7	1.7	147	11	Q9D0S2	Q9D0S2 mus musculus
43	7	1.7	150	6	Q9SG64	Q9SG64 oryctolagus
44	7	1.7	162	10	Q9EPD4	Q9EPD4 petunia int
45	7	1.7	163	11	Q9JLY9	Q9JLY9 mus musculus

#### ALIGNMENTS

RESULT 1	Q9UK79	PRELIMINARY;	PRT;	419 AA.
AC	Q9UK79;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	HERstatin.			
GN	HER-2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99415951; PubMed=10485918;			
RA	Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;			
RT	"The HER-2/neu Receptor Tyrosine Kinase gene encodes a secreted			
RT	autoinhibitor.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;			
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF177761; AAD56009.2; -			
DR	InterPro: IPR000494; EGFR_L.domain.			
DR	InterPro: IPR002174; Furin-like.			
DR	Pfam: PF00757; Furin-like; 1.			
DR	Pfam: PF01030; Recep_L.domain; 1.			
DR	SMART; SM00261; FU; 1.			
SQ	SEQUENCE 419 AA; 45472 MW; FECLIB347E2D030C CRC64;			
Query Match	82.3%; Score 345; DB 4; Length 419;			
Best Local Similarity	100.0%; Pred. NO. 0;			
Matches 345; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
Oy	1 MELALACRMGLLALLPPGAASVQCTGTDMLRLPASPETHLDMRLHYGCGVQGNL 60			
Db	1 MELALACRMGLLALLPPGAASVQCTGTDMLRLPASPETHLDMRLHYGCGVQGNL 60			

```

OY 61 ELTYLPTNASTSLFQDIOEVOGYVLLAHNQVROVPLQRLRIYRGTOLEFEDNVALAVDNG 120
    |||||||
DB 61 ELTYLPTNASTSLFQDIOEVOGYVLLAHNQVROVPLQRLRIYRGTOLEFEDNVALAVDNG 120
OY 121 DPLNNTPTVGTGASPGGLRELQRLSLREILKGVLIQRPOLCYQDTITLMDIFHKNNOQA 180
    |||||||
DB 121 DPLNNTPTVGTGASPGGLRELQRLSLREILKGVLIQRPOLCYQDTITLMDIFHKNNOQA 180
OY 181 LTLIDNRSRACHPCSPMGSGRCWESSSDCOSTRTVYAGGCARCKGPLPDDCCEQC 240
    |||||||
DB 181 LTLIDNRSRACHPCSPMGSGRCWESSSDCOSTRTVYAGGCARCKGPLPDDCCEQC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYTFGASCYTAC 300
    |||||||
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYTFGASCYTAC 300
OY 301 YNLTSTDVGSCTLVCPPLHNOVTAEDGTORCEKSKPCARGTSL 345
    |||||||
DB 301 YNLTSTDVGSCTLVCPPLHNOVTAEDGTORCEKSKPCARGTSL 345

```

## RESULT 2

```

OY 61 ELTYLPTNASTSLFQDIOEVOGYVLLAHNQVROVPLQRLRIYRGTOLEFEDNVALAVDNG 120
    |||||||
DB 61 ELTYLPTNASTSLFQDIOEVOGYVLLAHNQVROVPLQRLRIYRGTOLEFEDNVALAVDNG 120
OY 121 DPLNNTPTVGTGASPGGLRELQRLSLREILKGVLIQRPOLCYQDTITLMDIFHKNNOQA 180
    |||||||
DB 121 DPLNNTPTVGTGASPGGLRELQRLSLREILKGVLIQRPOLCYQDTITLMDIFHKNNOQA 180
OY 181 LTLIDNRSRACHPCSPMGSGRCWESSSDCOSTRTVYAGGCARCKGPLPDDCCEQC 240
    |||||||
DB 181 LTLIDNRSRACHPCSPMGSGRCWESSSDCOSTRTVYAGGCARCKGPLPDDCCEQC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYTFGASCYTAC 300
    |||||||
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYTFGASCYTAC 300
OY 301 YNLTSTDVGSCTLVCPPLHNOVTAEDGTORCEKSKPCARGTSL 345
    |||||||
DB 301 YNLTSTDVGSCTLVCPPLHNOVTAEDGTORCEKSKPCARGTSL 345

```

Query Match 16.0%; Score 67; DB 6; Length 149;

Best local Similarity 100.0%; Pred. No. 2e-60; Mismatches 0; Indels 0; Gaps 0;

```

OY 231 LPTDCHEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYT 290
    |||||||
DB 74 LPTDCHEQCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYT 133
OY 291 FGASCVT 297
    |||||||
DB 134 FGASCVT 140

```

## RESULT 3

```

OY 61 ELTYLPTNASTSLFQDIOEVOGYVLLAHNQVROVPLQRLRIYRGTOLEFEDNVALAVDNG 120
    |||||||
DB 61 ELTYLPTNASTSLFQDIOEVOGYVLLAHNQVROVPLQRLRIYRGTOLEFEDNVALAVDNG 120
OY 121 DPLNNTPTVGTGASPGGLRELQRLSLREILKGVLIQRPOLCYQDTITLMDIFHKNNOQA 180
    |||||||
DB 121 DPLNNTPTVGTGASPGGLRELQRLSLREILKGVLIQRPOLCYQDTITLMDIFHKNNOQA 180
OY 181 LTLIDNRSRACHPCSPMGSGRCWESSSDCOSTRTVYAGGCARCKGPLPDDCCEQC 240
    |||||||
DB 181 LTLIDNRSRACHPCSPMGSGRCWESSSDCOSTRTVYAGGCARCKGPLPDDCCEQC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYTFGASCYTAC 300
    |||||||
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYTFGASCYTAC 300
OY 301 YNLTSTDVGSCTLVCPPLHNOVTAEDGTORCEKSKPCARGTSL 345
    |||||||
DB 301 YNLTSTDVGSCTLVCPPLHNOVTAEDGTORCEKSKPCARGTSL 345

```

```

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbb-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB008451; BAA23127.1; -.
DR HSSP: P11362; 1FCG.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; Fy; 3.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364DA9CAACD46 CRC64;

```

Query Match 15.8%; Score 66; DB 6; Length 1259;

Best local Similarity 100.0%; Pred. No. 1.5e-58; Mismatches 0; Indels 0; Gaps 0;

```

OY 232 PTDCHQCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYT 291
    |||||||
DB 232 PTDCHQCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYT 291
OY 292 GASCVT 297
    |||||||
DB 292 GASCVT 297

```

## RESULT 4

```

OY 61 ELTYLPTNASTSLFQDIOEVOGYVLLAHNQVROVPLQRLRIYRGTOLEFEDNVALAVDNG 120
    |||||||
DB 61 ELTYLPTNASTSLFQDIOEVOGYVLLAHNQVROVPLQRLRIYRGTOLEFEDNVALAVDNG 120
OY 121 DPLNNTPTVGTGASPGGLRELQRLSLREILKGVLIQRPOLCYQDTITLMDIFHKNNOQA 180
    |||||||
DB 121 DPLNNTPTVGTGASPGGLRELQRLSLREILKGVLIQRPOLCYQDTITLMDIFHKNNOQA 180
OY 181 LTLIDNRSRACHPCSPMGSGRCWESSSDCOSTRTVYAGGCARCKGPLPDDCCEQC 240
    |||||||
DB 181 LTLIDNRSRACHPCSPMGSGRCWESSSDCOSTRTVYAGGCARCKGPLPDDCCEQC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYTFGASCYTAC 300
    |||||||
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRYTFGASCYTAC 300
OY 301 YNLTSTDVGSCTLVCPPLHNOVTAEDGTORCEKSKPCARGTSL 345
    |||||||
DB 301 YNLTSTDVGSCTLVCPPLHNOVTAEDGTORCEKSKPCARGTSL 345

```

Query Match 2.1%; Score 9; DB 11; Length 478;

Best local Similarity 100.0%; Pred. No. 2.4; Mismatches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 QCAAGCTGP 247  
 |||||||  
 DB 235 QCAAGCTGP 243

## RESULT 5

Q90836

AC 090836 PRELIMINARY: PRT: 527 AA.

DR 01-NOV-1996 (TREMBLrel. 01, Created)

DR 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE EGF/EGF-alpha receptor precursor.

GN C-ERBB.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=92123214; PubMed=1732751;

RA Flickinger T.W., Maibhe N.J., Kung H.-J.;

RT "An alternatively processed mRNA from the avian c-erbB gene encodes a

soluble, truncated form of the receptor that can block ligand-

dependent transformation."

RT Mol. Cell. Biol. 12:883-893(1992).

DR EMBL: M77637; AAA48759.1; -

DR InterPro: IPR000494; EGFR\_L\_domain.

DR InterPro: IPR002174; Furin-like.

DR Pfam: PF00757; Furin-like; 1.

DR Pfam: PF01030; Recep\_L\_domain; 2.

DR SMART: SM00261; FU; 2.

KW Receptor; Signal.

FT SIGNAL 1 28

FT CHAIN 29 527

FT SEQUENCE 527 AA; 58353 MW; 764564ABCC095298 CRC64;

Query Match

Best Local Similarity 2.1%; Score 9; DB 13; Length 527;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 QCAAGCTGP 247

|||||||

DB 245 QCAAGCTGP 253

## RESULT 6

Q9PSH2

AC 09PSH2 PRELIMINARY: PRT: 599 AA.

DR 01-MAY-2000 (TREMBLrel. 13, Created)

DR 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Epidermal growth factor receptor (Fragment).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=94020816; PubMed=8414496;

RA Callaghan T., Antczak M., Flickinger T., Raines M.,

Kung H.-J.;

RT "A complete description of the EGF-receptor exon structure;

RT implication in oncogenic activation and domain evolution."

RT Oncogene 8:2939-2948(1993).

DR InterPro: IPR000494; EGFR\_L\_domain.

DR InterPro: IPR002174; Furin-like.

DR Pfam: PF00757; Furin-like; 2.

DR Pfam: PF01030; Recep\_L\_domain; 2.

DR SMART: SM00261; FU; 3.  
 SQ SEQUENCE 599 AA; 66363 MW; FEAB46D23D991BD CRC64;

Query Match

Best Local Similarity 2.1%; Score 9; DB 13; Length 599;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 QCAAGCTGP 247

|||||||

DB 393 QCAAGCTGP 401

## RESULT 7

Q9ERV6

AC 09ERV6 PRELIMINARY: PRT: 643 AA.

DR 01-MAR-2001 (TREMBLrel. 16, Created)

DR 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Epidermal growth factor receptor isoform 2.

GN EGFR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV/TAC;

RA Reiter J.L., Threagill D.W., Eley G.D., Strunk K.E., Danielson A.J.,

RA Sinclair J.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,

RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,

Maibhe N.J.;

RT "Comparative genomic sequence analysis and isolation of human and

mouse alternative Egr1 transcripts encoding truncated receptor

isoforms."

RT Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF275366; AAG28046.1; JOINED.

DR EMBL: AF275364; AAG28046.1; JOINED.

DR EMBL: AF275365; AAG28046.1; JOINED.

DR MGD: MGI:95294; Egr1.

DR InterPro: IPR000345; Cytochrome\_b.

DR InterPro: IPR000494; EGFR\_L\_domain.

DR InterPro: IPR002174; Furin-like.

DR Pfam: PF00757; Furin-like; 1.

DR Pfam: PF01030; Recep\_L\_domain; 2.

DR SMART: SM00261; FU; 4.

DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.

KW Receptor.

SQ SEQUENCE 643 AA; 71476 MW; DEF22002C84911B1 CRC64;

Query Match

Best Local Similarity 2.1%; Score 9; DB 11; Length 643;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 QCAAGCTGP 247

|||||||

DB 235 QCAAGCTGP 243

## RESULT 8

Q9WVF5

AC 09WVF5 PRELIMINARY: PRT: 655 AA.

DR 01-NOV-1999 (TREMBLrel. 12, Created)

DR 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Epidermal growth factor receptor (Epidermal growth factor receptor

isoform 3).

GN EGFR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threaddgill D.W., Danielsen A.J., Schell C.,
RA Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RN Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threaddgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearshall R.S., Green P.J., Yee D., Lampard A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21083660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki H., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda Y.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl S., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AF124513; AAD4149.1; -
DR EMBL: AF275366; AAG28047.1; -
DR EMBL: AF275364; AAG28047.1; JOINED.
DR EMBL: AF275365; AAG28047.1; JOINED.
DR EMBL: AK004944; BAB23688.1; -
DR EMBL: AK004911; BAB23641.1; -
DR MGD: MGI:95294; Egfr.
DR InterPro: IPR000494; EGFR_L.domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L.domain; 2.
DR SMART: SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B340631BC928CB CRC64;

```

```

DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Epidermal growth factor receptor.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RX MEDLINE=90258888; PubMed=2342466;
RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue.";
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A.;
RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RX Gutridge K., Dawson T.L., Earp H.S.;
RA Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: M37394; AF144008.1; -
DR HSP: P11362; IFGK.
DR InterPro: IPR000494; EGFR_L.domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002145; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF01030; Recep_L.domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 3
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1209 AA; 134891 MW; 96FE7F6CC1B773 CRC64;

```

Query Match 2.1%; Score 9; DB 11; Length 1209;  
Best Local Similarity 100.0%; Pred. No. 5.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 239 OCAAGCTGP 247
DB 235 OCAAGCTGP 243

```

RESULT 10  
Q9EP98 PRELIMINARY; PRT; 1210 AA.  
AC Q9EP98.  
DT 01-MAR-2001 (TRENBLREL. 16, Created)  
DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)  
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
DE Epidermal growth factor receptor isoform 1.  
GN EGFR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;  
RA Reiter J.L., Threaddgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,



```

RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maine N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr1 transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Schenl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maine N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr1 transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275366; AAG28045.1; -.
DR EMBL; AF275364; AAG28045.1; JOINED.
DR EMBL; AF275365; AAG28045.1; JOINED.
DR EMBL; AF275367; AAG24386.1; -.
DR HSSP; P11362; 1FGK.
DR MGD; MGI:95294; Egfr.
DR InterPro: IPR000345; Cyto_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; EGFR_L_domain.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Receptor; transferase.
SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match
Best Local Similarity 2.1%; Score 9; DB 11; Length 1210;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 239 QCAAGCTGP 247
DB 235 QCAAGCTGP 243

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=91172172; PubMed=1672439;
RA Van D.H., Hung M.C.;
RT "Identification and characterization of a novel enhancer for the rat
RT neu promoter.";
RL Mol. Cell. Biol. 11:1875-1882(1991).
DR EMBL; M61004; AAA41686.1; -.
KW Oncogene; Transforming protein.
FT NON_TER
SQ SEQUENCE 48 AA; 4966 MW; 72A0E4B1E13F1129 CRC64;

Query Match
Best Local Similarity 1.9%; Score 8; DB 11; Length 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 LALLPPG 19
DB 15 LALLPPG 22

RESULT 12
O90216 PRELIMINARY; PRT; 63 AA.
AC O90216;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE Carbonyl reductase isoform I (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=OVARY;
RA Espey L.L., Yoshioaka S., Russell D., Ujioaka T., Vladu B., Skelsey M.,
RA Fujii S., Okamura H., Richards J.S.;
RT "Characterization of Ovarian Carbonyl Reductase Gene Expression during
RT Ovulation in the Gonadotropin-Primed Immature Rat.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF181955; AAF03394.1; -.
FT NON_TER
SQ SEQUENCE 63 AA; 6947 MW; 03514BC1F399AA56 CRC64;

Query Match
Best Local Similarity 1.9%; Score 8; DB 11; Length 63;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 LALLPPGA 20
DB 40 LALLPPGA 47

RESULT 13
O90215 PRELIMINARY; PRT; 63 AA.
AC O90215;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE Carbonyl reductase isoform II (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=OVARY;
RA Espey L.L., Yoshioaka S., Russell D., Ujioaka T., Vladu B., Skelsey M.,
RA Fujii S., Okamura H., Richards J.S.;
RT "Characterization of Ovarian Carbonyl Reductase Gene Expression during
RT Ovulation in the Gonadotropin-Primed Immature Rat.";

```

RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF181956; AAF03395.1; -  
 FT NON\_TER 1 1  
 SQ SEQUENCE 63 AA; 6916 MW; D5964BC1E299A2F4 CRC64;

Query Match 1.9%; Score 8; DB 11; Length 63;  
 Best Local Similarity 100.0%; Pred. No. 4.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 LALLPGA 20  
 |||||||  
 Db 40 LALLPGA 47

## RESULT 14

O9N6W6 PRELIMINARY; PRT; 217 AA.  
 AC O9N6W6;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE HLHgamma protein.  
 GN HLHM-GAMMA OR F.  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7224;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cooper M.T.D., Tyler D.M., Furiols M., Chalkiadaki A., Delidakis C.,  
 RA Bray S.;  
 RT "Spatially restricted factors co-operate with Notch in the regulation  
 of Enhancer of split genes.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ276313; CAB77019.1;  
 DR Flybase; FBgn0012370; Dhyd\HLhm-gamma.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR InterPro; IPR003650; Orange.  
 DR Pfam; PF00010; HLH; 1.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00511; ORANGE; 1.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
 SQ SEQUENCE 217 AA; 24082 MW; 39177C58F322B4ED CRC64;

Query Match 1.9%; Score 8; DB 5; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 392 SVRSPVS 399  
 |||||||  
 Db 159 SVRSPVS 166

## RESULT 15

O08558 PRELIMINARY; PRT; 277 AA.  
 AC O08558;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Inducible carbonyl reductase.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-MISTER; TISUE-OVARY;  
 RX MEDLINE=97167735; Pubmed=9015353;  
 RA Aoki H., Okada T., Mizutani T., Numata Y., Minegishi T., Miyamoto K.;  
 RT "Identification of two closely related genes, inducible and  
 noninducible carbonyl reductases in the rat ovary.";

RL Biochem. Biophys. Res. Commun. 230:518-523(1997).  
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 (SDR) FAMILY.  
 DR EMBL; D89069; BAAL9007.1; -  
 DR HSSP; P50162; IAE1.  
 DR InterPro; IPR002198; ADH\_short.  
 DR Pfam; PF00106; adh\_short; 1.  
 DR PROSITE; PS00061; ADH\_SHORT; UNKNOWN\_1.  
 KW Oxidoreductase.  
 SQ SEQUENCE 277 AA; 30654 MW; DDA015D1B71673A7 CRC64;

Query Match 1.9%; Score 8; DB 11; Length 277;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 LALLPGA 20  
 |||||||  
 Db 254 LALLPGA 261

Search completed: March 4, 2003, 12:46:40  
 Job time : 48.2751 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:43:52 ; Search time 33.6546 Seconds  
(without alignments)  
366.315 Million cell updates/sec

Title: US-09-234-208b-2

Sequence: 1 MEALALCRWGLLALLPPGA.....VGRGPDPAHVAVNLSRYEG 419

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents\_AA:\*

- 1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*
- 4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*
- 5: /cgn2\_6/prodata/1/1aa/PCUT05\_COMB.pep:\*
- 6: /cgn2\_6/prodata/1/1aa/backfillset1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	419	100.0	419	4	US-09-630-155-2	Sequence 2, Appl1
2	340	81.1	782	2	US-09-146-283-4	Sequence 4, Appl1
3	340	81.1	782	3	US-08-579-823A-4	Sequence 4, Appl1
4	340	81.1	782	4	US-09-344-195-4	Sequence 4, Appl1
5	340	81.1	1255	1	US-08-467-083-68	Sequence 68, Appl1
6	340	81.1	1255	1	US-08-414-417B-68	Sequence 68, Appl1
7	340	81.1	1255	2	US-08-484-438-8	Sequence 8, Appl1
8	340	81.1	1255	2	US-08-486-348A-68	Sequence 8, Appl1
9	340	81.1	1255	2	US-08-625-101-2	Sequence 2, Appl1
10	340	81.1	1255	2	US-08-468-545B-68	Sequence 68, Appl1
11	340	81.1	1255	3	US-08-356-786-2	Sequence 2, Appl1
12	340	81.1	1255	3	US-08-466-680B-68	Sequence 68, Appl1
13	319	76.1	624	3	US-08-422-108-1	Sequence 1, Appl1
14	319	76.1	624	4	US-08-422-734-1	Sequence 1, Appl1
15	96	22.9	97	4	US-09-046-783-3	Sequence 3, Appl1
16	96	22.9	97	4	US-09-046-783-3	Sequence 3, Appl1
17	79	18.9	79	4	US-09-630-155-1	Sequence 1, Appl1
18	17	4.1	17	1	US-08-467-083-61	Sequence 61, Appl1
19	17	4.1	17	1	US-08-414-417B-61	Sequence 61, Appl1
20	17	4.1	17	2	US-08-486-348A-61	Sequence 61, Appl1
21	17	4.1	17	2	US-08-468-545B-61	Sequence 61, Appl1
22	17	4.1	17	3	US-08-466-680B-61	Sequence 61, Appl1
23	15	3.6	15	1	US-08-467-083-30	Sequence 30, Appl1
24	15	3.6	15	1	US-08-467-083-31	Sequence 31, Appl1
25	15	3.6	15	1	US-08-467-083-32	Sequence 32, Appl1
26	15	3.6	15	1	US-08-467-083-33	Sequence 33, Appl1
27	15	3.6	15	1	US-08-467-083-56	Sequence 56, Appl1

28	15	3.6	15	1	US-08-414-417B-30	Sequence 30, Appl1
29	15	3.6	15	1	US-08-414-417B-31	Sequence 31, Appl1
30	15	3.6	15	1	US-08-414-417B-32	Sequence 32, Appl1
31	15	3.6	15	1	US-08-414-417B-33	Sequence 33, Appl1
32	15	3.6	15	1	US-08-414-417B-56	Sequence 56, Appl1
33	15	3.6	15	2	US-08-486-348A-30	Sequence 30, Appl1
34	15	3.6	15	2	US-08-486-348A-31	Sequence 31, Appl1
35	15	3.6	15	2	US-08-486-348A-32	Sequence 32, Appl1
36	15	3.6	15	2	US-08-486-348A-33	Sequence 33, Appl1
37	15	3.6	15	2	US-08-486-348A-56	Sequence 56, Appl1
38	15	3.6	15	2	US-08-468-545B-30	Sequence 30, Appl1
39	15	3.6	15	2	US-08-468-545B-31	Sequence 31, Appl1
40	15	3.6	15	2	US-08-468-545B-32	Sequence 32, Appl1
41	15	3.6	15	2	US-08-468-545B-33	Sequence 33, Appl1
42	15	3.6	15	2	US-08-468-545B-56	Sequence 56, Appl1
43	15	3.6	15	3	US-08-466-680B-30	Sequence 30, Appl1
44	15	3.6	15	3	US-08-466-680B-31	Sequence 31, Appl1
45	15	3.6	15	3	US-08-466-680B-32	Sequence 32, Appl1

ALIGNMENTS

RESULT 1  
US-09-630-155-2  
; Sequence 2, Application US/09630155  
; Patent No. 6414130  
GENERAL INFORMATION:  
; APPLICANT: Doherty, Joni Kristin and Gail M. Clinton  
; TITLE OF INVENTION: HER-2 BINDING ANTAGONISTS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVIS WRIGHT TREMAINE LLP  
; STREET: 1501 Fourth Avenue, 2600 Century Square  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98101  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: PC compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: word  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/630.155  
; FILING DATE: 16-Jan-2001  
; CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
; NAME: Davison, Barry L.  
; REGISTRATION NUMBER: 47,309  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206 628-7621  
; TELEFAX: 206 628-7699  
INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 419  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: polypeptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-630-155-2  
Query Match 100.0%; Score 419; DB 4; Length 419;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEALALCRWGLLALLPPGAASVQCTGTDMKRLRPASPTHLDMRLHYOGGVQGNL 60  
|||||  
Db 1 MEALALCRWGLLALLPPGAASVQCTGTDMKRLRPASPTHLDMRLHYOGGVQGNL 60  
QY 61 ELTYLPTNASTLSPLODIOEVGVVLIAHNOVROVPLQRLRIYVGTQLFEDNVALLAVLDNG 120

```
|||||
Db 61 ELTYLPTNLSLFLQDIQEVGYVLLAHNQVROVPLQRLRIYRGTLQDFEDNYALAVLDNG 120
QY 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180
Db 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180
QY 181 LTLIDNRSRACHPCSPMKGSRGWESSESDCSLRTVYAGGCARCKGPLPTDCCHEOC 240
Db 181 LTLIDNRSRACHPCSPMKGSRGWESSESDCSLRTVYAGGCARCKGPLPTDCCHEOC 240
QY 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
Db 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
QY 301 YNLTSTDVGSCTLVCPHLHNOEVTAEQDTCRCEKSRPCARHTSLPRPAVVPVLMOP 360
Db 301 YNLTSTDVGSCTLVCPHLHNOEVTAEQDTCRCEKSRPCARHTSLPRPAVVPVLMOP 360
QY 361 GPAHPVLSFLRSPMDLVSAFYSPLAPLSPVSPISVSGRGPDDPAHVAVNLSTREG 419
Db 361 GPAHPVLSFLRSPMDLVSAFYSPLAPLSPVSPISVSGRGPDDPAHVAVNLSTREG 419
```

RESULT 2  
US-09-146-283-4  
; Sequence 4, Application US/09146283  
; Patent No. 5976546

GENERAL INFORMATION:  
APPLICANT: Laus, Retner  
APPLICANT: Ruegg, Curtis L.  
APPLICANT: Wu, Hongyu  
TITLE OF INVENTION: Immunostimulatory Compositions  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Ave. Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/146,283  
FILING DATE: 03-SEPT-1998  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Judge, Linda R.  
REGISTRATION NUMBER: 42,702  
REFERENCE/DOCKET NUMBER: 7636-0010.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-324-0880  
TELEFAX: 650-324-0960  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 782 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: homo sapiens  
INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8  
US-09-146-283-4

Query Match 81.1%; Score 340; DB 2; Length 782;  
Best Local Similarity 100.0%; Pred. No. 9.7e-317;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 METALCRWGLLALLPPAASSTOVCTGDMKRLPASPETHLMDLRHLXYGCGQVQGNL 60
Db 1 METALCRWGLLALLPPAASSTOVCTGDMKRLPASPETHLMDLRHLXYGCGQVQGNL 60
QY 61 ELTYLPTNLSLFLQDIQEVGYVLLAHNQVROVPLQRLRIYRGTLQDFEDNYALAVLDNG 120
Db 61 ELTYLPTNLSLFLQDIQEVGYVLLAHNQVROVPLQRLRIYRGTLQDFEDNYALAVLDNG 120
QY 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180
Db 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180
QY 181 LTLIDNRSRACHPCSPMKGSRGWESSESDCSLRTVYAGGCARCKGPLPTDCCHEOC 240
Db 181 LTLIDNRSRACHPCSPMKGSRGWESSESDCSLRTVYAGGCARCKGPLPTDCCHEOC 240
QY 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
Db 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCYTACP 300
QY 301 YNLTSTDVGSCTLVCPHLHNOEVTAEQDTCRCEKSRPCAR 340
Db 301 YNLTSTDVGSCTLVCPHLHNOEVTAEQDTCRCEKSRPCAR 340
```

RESULT 3  
US-08-579-823A-4  
; Sequence 4, Application US/08579823A  
; Patent No. 6080409

GENERAL INFORMATION:  
APPLICANT: Laus, Retner  
APPLICANT: Ruegg, Curtis L.  
APPLICANT: Wu, Hongyu  
TITLE OF INVENTION: Immunostimulatory Composition and Method  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Ave. Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/579,823A  
FILING DATE: 03-DEC-1998  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Judge, Linda R.  
REGISTRATION NUMBER: 42,702  
REFERENCE/DOCKET NUMBER: 7636-0010  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-324-0880  
TELEFAX: 650-324-0960  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 782 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: homo sapiens  
INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8  
US-08-579-823A-4

Query Match 81.1%; Score 340; DB 3; Length 782;  
Best Local Similarity 100.0%; Pred. No. 9.7e-317;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEALALCRWGLLALLPPGAASOVCTGTDMLRLPASPTHLMRLHLYOGCOVVOGNTL 60  
Db 1 MEALALCRWGLLALLPPGAASOVCTGTDMLRLPASPTHLMRLHLYOGCOVVOGNTL 60  
QY 61 ELTYLPTNASTLFDIOIEOVGYVLLAHNOVROPLORLIVGTQOLFEDNVALAVLDNG 120  
Db 61 ELTYLPTNASTLFDIOIEOVGYVLLAHNOVROPLORLIVGTQOLFEDNVALAVLDNG 120  
QY 121 DPLNNTPTVAGSPGGLREQLRSLEFIILKGVLIOBNPOLCYODTILMKDIFHKNNOLA 180  
Db 121 DPLNNTPTVAGSPGGLREQLRSLEFIILKGVLIOBNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVCAGGCARCKGFLPTDCHEQC 240  
Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVCAGGCARCKGFLPTDCHEQC 240  
QY 241 AACCTGPKHSDCLACHFNHSGICEHLCPALVYNTDTEFSMPNPEGRTYFGASCVYACP 300  
Db 241 AACCTGPKHSDCLACHFNHSGICEHLCPALVYNTDTEFSMPNPEGRTYFGASCVYACP 300  
QY 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCAR 340  
Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCAR 340

## RESULT 4

US-09-344-195-4

; Sequence 4, Application US/09344195  
; Patent No. 6210662

GENERAL INFORMATION:

APPLICANT: Laus, Retner  
Ruegg, Curtlis L.

Mu, Hongyu

TITLE OF INVENTION: Immunostimulatory Compositions

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger &amp; Associates

STREET: 350 Cambridge Ave. Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/344,195

FILING DATE: 24-Jun-1999

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/146,283

FILING DATE: 03-SEPT-1998

ATTORNEY/AGENT INFORMATION:

NAME: Judge, Linda R.

REGISTRATION NUMBER: 42,702

REFERENCE/DOCKET NUMBER: 7636-0010.21

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-324-0880

TELEFAX: 650-324-0960

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 782 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: homo sapiens

INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-344-195-4

Query Match 81.1%; Score 340; DB 4; Length 782;

Best Local Similarity 100.0%; Pred No. 9.7e-317; Indels 0; Gaps 0;

Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEALALCRWGLLALLPPGAASOVCTGTDMLRLPASPTHLMRLHLYOGCOVVOGNTL 60  
Db 1 MEALALCRWGLLALLPPGAASOVCTGTDMLRLPASPTHLMRLHLYOGCOVVOGNTL 60  
QY 61 ELTYLPTNASTLFDIOIEOVGYVLLAHNOVROPLORLIVGTQOLFEDNVALAVLDNG 120  
Db 61 ELTYLPTNASTLFDIOIEOVGYVLLAHNOVROPLORLIVGTQOLFEDNVALAVLDNG 120  
QY 121 DPLNNTPTVAGSPGGLREQLRSLEFIILKGVLIOBNPOLCYODTILMKDIFHKNNOLA 180  
Db 121 DPLNNTPTVAGSPGGLREQLRSLEFIILKGVLIOBNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVCAGGCARCKGFLPTDCHEQC 240  
Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVCAGGCARCKGFLPTDCHEQC 240  
QY 241 AACCTGPKHSDCLACHFNHSGICEHLCPALVYNTDTEFSMPNPEGRTYFGASCVYACP 300  
Db 241 AACCTGPKHSDCLACHFNHSGICEHLCPALVYNTDTEFSMPNPEGRTYFGASCVYACP 300  
QY 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCAR 340  
Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCAR 340

## RESULT 5

US-08-467-083-68

; Sequence 68, Application US/08467083  
; Patent No. 5726023

GENERAL INFORMATION:

APPLICANT: Cheever, Martin A.

TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN

TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE

NUMBER OF SEQUENCES: 68

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed and Berry

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: US

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,083

FILING DATE: 06-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/414,417

FILING DATE: 06-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Sharkey, Richard G.

REGISTRATION NUMBER: 32,629

REFERENCE/DOCKET NUMBER: 920010.448C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

TELEX: 3723836 SEEDANBERY

INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:

LENGTH: 1255 amino acids

TYPE: amino acid

TOPOLOGY: linear  
US-08-467-083-68

Query Match 81.1%; Score 340; DB 1; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 1.5e-316;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASVQVCTGTDMLRLPASFETHDMLRHLXGCGVQVGNL 60  
DB 1 MELAALCRWGLLALLPPGAASVQVCTGTDMLRLPASFETHDMLRHLXGCGVQVGNL 60  
QY 61 ELTYLPTNLSLFDIOIEVQGVYLLAHNOVROVPLQRLRIYRGTDLFEDNVALAVLDNG 120  
DB 61 ELTYLPTNLSLFDIOIEVQGVYLLAHNOVROVPLQRLRIYRGTDLFEDNVALAVLDNG 120  
QY 121 DPLNNTPTVGTASPGGLREQLRLSTLEILKGVLLIORNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVGTASPGGLREQLRLSTLEILKGVLLIORNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRVCAGCARGKGPLPDCCHQC 240  
DB 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRVCAGCARGKGPLPDCCHQC 240  
QY 241 AACCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFGASCVTACP 300  
DB 241 AACCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPDLHNOEVTAEEDGTORCEKSKPCAR 340  
DB 301 YNYLSTDVGSCTLVCPDLHNOEVTAEEDGTORCEKSKPCAR 340

RESULT 6  
US-08-414-417B-68

Sequence 68, Application US/08414417B

Patent No. 5801005

GENERAL INFORMATION:

APPLICANT: Cheever, Martin A.

APPLICANT: Disis, Mary L.

TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN

TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE

TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Seed and Berry LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: US

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/414,417B

FILING DATE: 31-MAR-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Sharkey, Richard G.

REGISTRATION NUMBER: 32,629

REFERENCE/DOCKET NUMBER: 920010.448C2

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:

LENGTH: 1255 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-414-417B-68

Query Match 81.1%; Score 340; DB 1; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 1.5e-316;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASVQVCTGTDMLRLPASFETHDMLRHLXGCGVQVGNL 60  
DB 1 MELAALCRWGLLALLPPGAASVQVCTGTDMLRLPASFETHDMLRHLXGCGVQVGNL 60  
QY 61 ELTYLPTNLSLFDIOIEVQGVYLLAHNOVROVPLQRLRIYRGTDLFEDNVALAVLDNG 120  
DB 61 ELTYLPTNLSLFDIOIEVQGVYLLAHNOVROVPLQRLRIYRGTDLFEDNVALAVLDNG 120  
QY 121 DPLNNTPTVGTASPGGLREQLRLSTLEILKGVLLIORNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVGTASPGGLREQLRLSTLEILKGVLLIORNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRVCAGCARGKGPLPDCCHQC 240  
DB 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRVCAGCARGKGPLPDCCHQC 240  
QY 241 AACCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFGASCVTACP 300  
DB 241 AACCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPDLHNOEVTAEEDGTORCEKSKPCAR 340  
DB 301 YNYLSTDVGSCTLVCPDLHNOEVTAEEDGTORCEKSKPCAR 340

RESULT 7  
US-08-484-438-8

Sequence 8, Application US/08484438

Patent No. 5811098

Patent No. 5811098 5780031

GENERAL INFORMATION:

APPLICANT: Plovman, Gregory D.

APPLICANT: Culouscou, Jean-Michel

APPLICANT: Shoyab, Mohammed

APPLICANT: Siegall, Clay B.

APPLICANT: Helster m, Ingegerd

APPLICANT: Helster m, Karl E.

TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE

TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,438

FILING DATE: 07-JUN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/323,442

FILING DATE: 14-OCT-1994

APPLICATION NUMBER: US 08/150,704

FILING DATE: 10-NOV-1993

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/981,165

FILING DATE: 24-NOV-1992

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Mistock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 5624-230

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-484-438-8

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e-316; Length 1255;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPRGAASFOVCTGTDMLRLPASPTHLDMLRHLYQGCGVQVGNL 60
DB 1 MELAALCRWGLLALLPRGAASFOVCTGTDMLRLPASPTHLDMLRHLYQGCGVQVGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVVLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVVLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLIORNPOLCYQDPTLMKDFHKNQOLA 180
DB 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLIORNPOLCYQDPTLMKDFHKNQOLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGRCWSESSDCSLTRTVAGGACARCKGPLPDDCHEQC 240
DB 181 LTLIDNRSRACHPCSPMCKSGRCWSESSDCSLTRTVAGGACARCKGPLPDDCHEQC 240
QY 241 AACGTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMNPBGRYFFGASCVYACP 300
DB 241 AACGTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMNPBGRYFFGASCVYACP 300
QY 301 YNYLSTDVGSCTLVCPHLHNOEVAEDGTORCEKSKPCAR 340
DB 301 YNYLSTDVGSCTLVCPHLHNOEVAEDGTORCEKSKPCAR 340

RESULT 8
US-08-486-348A-68
; Sequence 68, Application US/08486348A
; Patent No. 5846538
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,348A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C6
```

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 68
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-486-348A-68

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e-316; Length 1255;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPRGAASFOVCTGTDMLRLPASPTHLDMLRHLYQGCGVQVGNL 60
DB 1 MELAALCRWGLLALLPRGAASFOVCTGTDMLRLPASPTHLDMLRHLYQGCGVQVGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVVLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVVLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLIORNPOLCYQDPTLMKDFHKNQOLA 180
DB 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLIORNPOLCYQDPTLMKDFHKNQOLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGRCWSESSDCSLTRTVAGGACARCKGPLPDDCHEQC 240
DB 181 LTLIDNRSRACHPCSPMCKSGRCWSESSDCSLTRTVAGGACARCKGPLPDDCHEQC 240
QY 241 AACGTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMNPBGRYFFGASCVYACP 300
DB 241 AACGTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMNPBGRYFFGASCVYACP 300
QY 301 YNYLSTDVGSCTLVCPHLHNOEVAEDGTORCEKSKPCAR 340
DB 301 YNYLSTDVGSCTLVCPHLHNOEVAEDGTORCEKSKPCAR 340

RESULT 9
US-08-625-101-2
; Sequence 2, Application US/08625101
; Patent No. 5869445
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE
; TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION
; TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu
; TITLE OF INVENTION: ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,101
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
```

TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-625-101-2

Query Match 81.1%; Score 340; DB 2; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 1.5e-316;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASVQVCTGTDMLRLPASPEHLDMLRHLYQGCQVQVGNL 60  
DB 1 MELAALCRWGLLALLPPGAASVQVCTGTDMLRLPASPEHLDMLRHLYQGCQVQVGNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIVGTQLFEDNVALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIVGTQLFEDNVALAVLDNG 120  
QY 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLRTVCAGGCARCKGPLPTDCHEQC 240  
DB 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLRTVCAGGCARCKGPLPTDCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCVTACP 300  
QY 301 YNLTSDVGSCTLVCPLHNOEVAEDGTORCEKCSKPCAR 340  
DB 301 YNLTSDVGSCTLVCPLHNOEVAEDGTORCEKCSKPCAR 340

## RESULT 10

US-08-468-545B-68  
Sequence 68, Application US/08468545B  
Patent No. 5876712  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Disis, Mary L.  
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,545B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Shaikhey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-468-545B-68

Query Match 81.1%; Score 340; DB 2; Length 1255;  
Best Local Similarity 100.0%; Pred. No. 1.5e-316;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASVQVCTGTDMLRLPASPEHLDMLRHLYQGCQVQVGNL 60  
DB 1 MELAALCRWGLLALLPPGAASVQVCTGTDMLRLPASPEHLDMLRHLYQGCQVQVGNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIVGTQLFEDNVALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLAHNOVROVPLQRLRIVGTQLFEDNVALAVLDNG 120  
QY 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLRTVCAGGCARCKGPLPTDCHEQC 240  
DB 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLRTVCAGGCARCKGPLPTDCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCVTACP 300  
QY 301 YNLTSDVGSCTLVCPLHNOEVAEDGTORCEKCSKPCAR 340  
DB 301 YNLTSDVGSCTLVCPLHNOEVAEDGTORCEKCSKPCAR 340

## RESULT 11

US-08-356-786-2  
Sequence 2, Application US/08356786  
Patent No. 5877305  
GENERAL INFORMATION:  
APPLICANT: Huston, James S.  
APPLICANT: Oppermann, Hermann  
APPLICANT: Houston, L. L.  
APPLICANT: Ring, David B.  
TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer  
TITLE OF INVENTION: Marker  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault  
STREET: Exchange Place, 53 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,786  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/831,967  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pitcher, Edmund R.  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: CRP-053  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100



; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1255 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-356-786-2

Query Match 81.1%; Score 340; DB 2; Length 1255;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-316;  
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASVQCTGDMKRLRPASPEHLDMLRHLYQCCQVQGNL 60  
 DB 1 MELAALCRWGLLALLPPGAASVQCTGDMKRLRPASPEHLDMLRHLYQCCQVQGNL 60  
 QY 61 ELTYLPTNASLSFLDIOEVQGVVLAHNOVROPRLQRLRIYRGTOLEFEDNALAVLNDG 120  
 DB 61 ELTYLPTNASLSFLDIOEVQGVVLAHNOVROPRLQRLRIYRGTOLEFEDNALAVLNDG 120  
 QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILMKDIFHKNNOLA 180  
 DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILMKDIFHKNNOLA 180  
 QY 181 LFLIDNRSRACHPCSPMKGSRGWESSEDCQSLRTVACAGGCARCKGPLPTDCCHEQC 240  
 DB 181 LFLIDNRSRACHPCSPMKGSRGWESSEDCQSLRTVACAGGCARCKGPLPTDCCHEQC 240  
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTYFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTYFGASCVTACP 300  
 QY 301 YNVLSTDVGSCTLVCPPLHNOEVTAEEDGTORCEKCSKPCAR 340  
 DB 301 YNVLSTDVGSCTLVCPPLHNOEVTAEEDGTORCEKCSKPCAR 340

# RESULT 12

; US-08-466-680B-68  
 ; Sequence 68, Application US/08466680B  
 ; Patent No. 6075122

; GENERAL INFORMATION:  
 ; APPLICANT: Cheever, Martin A.  
 ; APPLICANT: Disis, Mary L.  
 ; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
 ; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
 ; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED  
 ; NUMBER OF SEQUENCES: 69  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Seed and Berry LLP  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: US  
 ; ZIP: 98104-7092  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/466,680B  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Sharkey, Richard G.  
 ; REGISTRATION NUMBER: 32,629  
 ; REFERENCE/DOCKET NUMBER: 920010.448C4  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 68:  
 ; SEQUENCE CHARACTERISTICS:

; LENGTH: 1255 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; US-08-466-680B-68

Query Match 81.1%; Score 340; DB 3; Length 1255;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-316;  
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASVQCTGDMKRLRPASPEHLDMLRHLYQCCQVQGNL 60  
 DB 1 MELAALCRWGLLALLPPGAASVQCTGDMKRLRPASPEHLDMLRHLYQCCQVQGNL 60  
 QY 61 ELTYLPTNASLSFLDIOEVQGVVLAHNOVROPRLQRLRIYRGTOLEFEDNALAVLNDG 120  
 DB 61 ELTYLPTNASLSFLDIOEVQGVVLAHNOVROPRLQRLRIYRGTOLEFEDNALAVLNDG 120  
 QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILMKDIFHKNNOLA 180  
 DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILMKDIFHKNNOLA 180  
 QY 181 LFLIDNRSRACHPCSPMKGSRGWESSEDCQSLRTVACAGGCARCKGPLPTDCCHEQC 240  
 DB 181 LFLIDNRSRACHPCSPMKGSRGWESSEDCQSLRTVACAGGCARCKGPLPTDCCHEQC 240  
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTYFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTYFGASCVTACP 300  
 QY 301 YNVLSTDVGSCTLVCPPLHNOEVTAEEDGTORCEKCSKPCAR 340  
 DB 301 YNVLSTDVGSCTLVCPPLHNOEVTAEEDGTORCEKCSKPCAR 340

# RESULT 13

; US-08-422-108-1  
 ; Sequence 1, Application US/08422108  
 ; Patent No. 6015567

; GENERAL INFORMATION:  
 ; APPLICANT: Hudziak, Robert M.  
 ; APPLICANT: Shepard, H. Michael  
 ; APPLICANT: Ullrich, Axel  
 ; TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Winpatin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/422,108  
 ; FILING DATE: 14-Apr-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/355460  
 ; FILING DATE: 13-DEC-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/048346  
 ; FILING DATE: 15-APR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/354319  
 ; FILING DATE: 19-MAY-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M  
 ; REGISTRATION NUMBER: 00,000  
 ; REFERENCE/DOCKET NUMBER: 554C2D2

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 624 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-422-108-1

Query Match
Best Local Similarity 76.1%; Score 319; DB 3; Length 624;
Best Local Similarity 100.0%; Pred. No. 9.4e-297;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 STQVCTGDMKRLPASPETHLDMRLHYQGCVVGNLELYLPTNLSLFLQDIQEVQ 81
DB 1 STQVCTGDMKRLPASPETHLDMRLHYQGCVVGNLELYLPTNLSLFLQDIQEVQ 60
QY 82 GYVLIANOVROVPLQRLRIVRGTOLEFEDNYALAVLDNGDPLNNTPTVYGASPGGLREIQ 141
DB 61 GYVLIANOVROVPLQRLRIVRGTOLEFEDNYALAVLDNGDPLNNTPTVYGASPGGLREIQ 120
QY 142 LRSLEIILKGVLIQIRNPOLCYODTILMKDIFHKNNQALTLIDTNRSRACHPCSPMCKG 201
DB 121 LRSLEIILKGVLIQIRNPOLCYODTILMKDIFHKNNQALTLIDTNRSRACHPCSPMCKG 180
QY 202 SRQWGSSEDCOSLTRVYAGGACARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHS 261
DB 181 SRQWGSSEDCOSLTRVYAGGACARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHS 240
QY 262 GICEIHCPLALVYNTDTEFSMPNPEGRTYFGASCYACPYNYLSDVGSCTLVCPHLNQE 321
DB 241 GICEIHCPLALVYNTDTEFSMPNPEGRTYFGASCYACPYNYLSDVGSCTLVCPHLNQE 300
QY 322 VTAEDGTORCEKSKPCAR 340
DB 301 VTAEDGTORCEKSKPCAR 319

RESULT 14
US-08-422-734-1
Sequence 1, Application US/08422734
Patent No. 6333169
GENERAL INFORMATION:
APPLICANT: Hudziak, Robert M.
APPLICANT: Shepard, H. Michael
APPLICANT: Ullrich, Axel
TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/422,734
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/422108
FILING DATE: 14-Apr-1995
APPLICATION NUMBER: 08/355460
FILING DATE: 13-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/048346
```

```
FILING DATE: 15-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/354319
FILING DATE: 19-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 554C2D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 624 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-422-734-1

Query Match
Best Local Similarity 76.1%; Score 319; DB 4; Length 624;
Best Local Similarity 100.0%; Pred. No. 9.4e-297;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 STQVCTGDMKRLPASPETHLDMRLHYQGCVVGNLELYLPTNLSLFLQDIQEVQ 81
DB 1 STQVCTGDMKRLPASPETHLDMRLHYQGCVVGNLELYLPTNLSLFLQDIQEVQ 60
QY 82 GYVLIANOVROVPLQRLRIVRGTOLEFEDNYALAVLDNGDPLNNTPTVYGASPGGLREIQ 141
DB 61 GYVLIANOVROVPLQRLRIVRGTOLEFEDNYALAVLDNGDPLNNTPTVYGASPGGLREIQ 120
QY 142 LRSLEIILKGVLIQIRNPOLCYODTILMKDIFHKNNQALTLIDTNRSRACHPCSPMCKG 201
DB 121 LRSLEIILKGVLIQIRNPOLCYODTILMKDIFHKNNQALTLIDTNRSRACHPCSPMCKG 180
QY 202 SRQWGSSEDCOSLTRVYAGGACARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHS 261
DB 181 SRQWGSSEDCOSLTRVYAGGACARCKGPLPTDCCHEQCAAGCTGPKHSDCLACHFNHS 240
QY 262 GICEIHCPLALVYNTDTEFSMPNPEGRTYFGASCYACPYNYLSDVGSCTLVCPHLNQE 321
DB 241 GICEIHCPLALVYNTDTEFSMPNPEGRTYFGASCYACPYNYLSDVGSCTLVCPHLNQE 300
QY 322 VTAEDGTORCEKSKPCAR 340
DB 301 VTAEDGTORCEKSKPCAR 319

RESULT 15
US-08-421-356-3
Sequence 3, Application US/08421356
Patent No. 5783404
GENERAL INFORMATION:
APPLICANT: Koski, Raymond A.
APPLICANT: Koski, Raymond A.
TITLE OF INVENTION: HER-2
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavenland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/421,356
FILING DATE:
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
```

NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-327  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 97 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-421-356-3

Query Match 22.9%; Score 96; DB 1; Length 97;  
Best Local Similarity 100.0%; Pred. No. 2.9e-84;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 LQRLRIVRGTOLEEDNYALAVLDNGDPLNNTTPVTGASPGGLRELOLRSLTEILKGVLI 155  
|||||  
DB 2 LQRLRIVRGTOLEEDNYALAVLDNGDPLNNTTPVTGASPGGLRELOLRSLTEILKGVLI 61  
QY 156 QRNPOLCYODTILMKDIFHKNNQLATLIDTNRRA 191  
|||||  
DB 62 QRNPOLCYODTILMKDIFHKNNQLATLIDTNRRA 97

Search completed: March 4, 2003, 12:48:29  
Job time : 35.6546 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 12:46:47 ; Search time 17.6687 Seconds  
(without alignments)  
1000.035 Million cell updates/sec

Title: US-09-234-208B-2

Perfect score: 419

Sequence: 1 METALCRWGLLALLPPGA.....VGRGPDPAHVAVNLRYEG 419

Scoring table: OLIGO

Gapop 60.0 , Gapect 60.0

Searched: 188354 seqs, 42170167 residues

Word size : 0

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCr\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCrUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	340	81.1	645	10 US-09-921-161-1	Sequence 1, Appli
2	340	81.1	653	9 US-09-854-356-3	Sequence 3, Appli
3	340	81.1	712	9 US-09-854-356-7	Sequence 7, Appli
4	340	81.1	919	9 US-09-854-356-6	Sequence 6, Appli
5	340	81.1	1235	9 US-09-769-508-2	Sequence 2, Appli
6	340	81.1	1255	9 US-09-854-356-1	Sequence 1, Appli
7	340	81.1	1255	9 US-09-930-125-2	Sequence 2, Appli
8	340	81.1	1255	9 US-09-441-411-6	Sequence 6, Appli
9	340	81.1	1255	10 US-09-811-123-9	Sequence 9, Appli
10	340	81.1	1255	10 US-09-811-115-3	Sequence 9, Appli
11	292	69.7	479	10 US-09-821-883-5	Sequence 5, Appli
12	292	69.7	555	10 US-09-821-883-1	Sequence 1, Appli
13	292	69.7	564	10 US-09-821-883-3	Sequence 3, Appli
14	292	69.7	690	10 US-09-821-883-2	Sequence 2, Appli
15	292	69.7	697	10 US-09-821-883-4	Sequence 4, Appli
16	289	66.0	289	10 US-09-821-883-4	Sequence 23, Appli
17	191	45.6	191	9 US-09-441-411-9	Sequence 9, Appli
18	52	12.4	654	9 US-09-854-356-8	Sequence 8, Appli
19	52	12.4	1256	9 US-09-854-356-2	Sequence 2, Appli

20	52	12.4	1260	9 US-09-870-759-118	Sequence 118, App
21	44	10.5	1256	9 US-09-854-356-14	Sequence 14, Appl
22	22	5.3	22	10 US-09-466-320-19	Sequence 19, Appl
23	22	5.3	68	10 US-09-466-320-11	Sequence 11, Appl
24	18	4.3	19	10 US-09-466-320-20	Sequence 20, Appl
25	18	4.3	65	10 US-09-466-320-12	Sequence 12, Appl
26	15	3.6	15	10 US-09-888-721-21	Sequence 21, Appl
27	14	3.3	14	9 US-10-001-546-31	Sequence 31, Appl
28	10	2.4	10	9 US-10-001-546-18	Sequence 30, Appl
29	9	2.1	9	9 US-09-017-743C-70	Sequence 70, Appl
30	9	2.1	9	9 US-10-001-546-17	Sequence 17, Appl
31	9	2.1	9	9 US-10-001-546-20	Sequence 20, Appl
32	9	2.1	9	9 US-10-001-546-21	Sequence 21, Appl
33	9	2.1	9	9 US-10-001-546-22	Sequence 22, Appl
34	9	2.1	478	10 US-09-867-521-2	Sequence 2, Appli
35	9	2.1	1210	10 US-09-725-433-2	Sequence 2, Appli
36	8	1.9	8	9 US-09-017-743C-92	Sequence 92, Appl
37	8	1.9	9	9 US-09-017-743C-105	Sequence 105, App
38	8	1.9	9	9 US-10-001-546-19	Sequence 19, Appl
39	8	1.9	10	9 US-10-001-546-18	Sequence 18, Appl
40	8	1.9	375	10 US-09-815-242-13383	Sequence 13383, A
41	8	1.9	375	10 US-09-815-242-13700	Sequence 13700, A
42	8	1.9	602	10 US-09-815-242-12896	Sequence 12896, A
43	7	1.7	35	9 US-09-981-876-204	Sequence 204, App
44	7	1.7	35	9 US-09-148-545-204	Sequence 204, App
45	7	1.7	36	10 US-09-864-761-48790	Sequence 48790, A

## ALIGNMENTS

```
RESULT 1
US-09-921-161-1
; Sequence 1, Application US/09921161
; Patent No. US2002009062A1
; GENERAL INFORMATION:
; APPLICANT: Ralphi, Peter
; TITLE OF INVENTION: ANALYTICAL METHOD
; FILE REFERENCE: GENENT 066A US/09/921,161
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/225,433
; PRIOR FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 645
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-921-161-1

Query Match      81.1%; Score 340; DB 10; Length 645;
Best Local Similarity 100.0%; Pred. No. 9.4e-303;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 METALCRWGLLALLPPGAASVOVCTGDMKILRLPASRETHIDMLRHLYOCQVQGNL 60
|||||
DB 1 METALCRWGLLALLPPGAASVOVCTGDMKILRLPASRETHIDMLRHLYOCQVQGNL 60
|||||
OY ELTYLPTNLSFLDIQEVQGVYLAHQVROVPLQRIYRGTOLEFDNYALAVLNG 120
|||||
DB ELTYLPTNLSFLDIQEVQGVYLAHQVROVPLQRIYRGTOLEFDNYALAVLNG 120
|||||
OY 61 ELTYLPTNLSFLDIQEVQGVYLAHQVROVPLQRIYRGTOLEFDNYALAVLNG 120
|||||
DB 61 ELTYLPTNLSFLDIQEVQGVYLAHQVROVPLQRIYRGTOLEFDNYALAVLNG 120
|||||
OY 121 DPLNNTPTVGTASPGGLRELQRLSLEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
|||||
DB 121 DPLNNTPTVGTASPGGLRELQRLSLEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180
|||||
OY 181 LTLIDNRSRACHPCSPMKSGRCWSESESDQSLTRYCAGGACRCKPLPTDCCHQC 240
|||||
DB 181 LTLIDNRSRACHPCSPMKSGRCWSESESDQSLTRYCAGGACRCKPLPTDCCHQC 240
|||||
OY 241 AAGCTGPKHSDCLACIHFHNSGICELCHPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
|||||
```

Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDFESHPNPEGRTYFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR 340  
Db 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR 340

RESULT 2  
US-09-854-356-3

; Sequence 3, Application US/09854356  
; Patent No. US20020177567A1  
; GENERAL INFORMATION:  
; APPLICANT: Cheever, Martin A.  
; APPLICANT: Gheysen, Dirk  
; APPLICANT: Corixa Corporation  
; APPLICANT: SmithKline Beecham Biologicals S. A.  
; TITLE OF INVENTION: HER-2/neu Fusion Proteins  
; FILE REFERENCE: 014058-009810PC  
; CURRENT APPLICATION NUMBER: US/09/854,356  
; CURRENT FILING DATE: 2001-05-09  
; PRIOR APPLICATION NUMBER: US 09/493,480  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: US 60/117,976  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 3  
; LENGTH: 653  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: extracellular domain (ECD) of human HER-2/neu  
US-09-854-356-3

Query Match 81.1%; Score 340; DB 9; Length 653;  
Best Local Similarity 100.0%; Pred. No. 9.5e-303;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASYOCTGTDMLRLPASPETHLDMRLHLYQGCQVVGNL 60  
Db 1 MELAALCRWGLLLALLPPGAASYOCTGTDMLRLPASPETHLDMRLHLYQGCQVVGNL 60  
QY 61 ELTYLPTNASLFLDIOGVGVYVLAHNOVQVPLQRLRIYRGQLFEDNVALAVLNG 120  
Db 61 ELTYLPTNASLFLDIOGVGVYVLAHNOVQVPLQRLRIYRGQLFEDNVALAVLNG 120  
QY 121 DPLNNTPTVTGASPGGLRLQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNOLA 180  
Db 121 DPLNNTPTVTGASPGGLRLQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNOLA 180  
QY 181 LFLIDTNRSRACHPCSPMKGSRGWESSEDCQSLRTVACAGCARKGRLPTDCCHEQC 240  
Db 181 LFLIDTNRSRACHPCSPMKGSRGWESSEDCQSLRTVACAGCARKGRLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDFESHPNPEGRTYFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDFESHPNPEGRTYFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR 340  
Db 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR 340

## RESULT 3

US-09-854-356-7  
; Sequence 7, Application US/09854356  
; Patent No. US20020177567A1  
; GENERAL INFORMATION:  
; APPLICANT: Cheever, Martin A.  
; APPLICANT: Gheysen, Dirk  
; APPLICANT: Corixa Corporation  
; APPLICANT: SmithKline Beecham Biologicals S. A.  
; TITLE OF INVENTION: HER-2/neu Fusion Proteins

; FILE REFERENCE: 014058-009810PC  
; CURRENT APPLICATION NUMBER: US/09/854,356  
; CURRENT FILING DATE: 2001-05-09  
; PRIOR APPLICATION NUMBER: US 09/493,480  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: US 60/117,976  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 7  
; LENGTH: 712  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein  
US-09-854-356-7

Query Match 81.1%; Score 340; DB 9; Length 712;  
Best Local Similarity 100.0%; Pred. No. 1e-302;  
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASYOCTGTDMLRLPASPETHLDMRLHLYQGCQVVGNL 60  
Db 1 MELAALCRWGLLLALLPPGAASYOCTGTDMLRLPASPETHLDMRLHLYQGCQVVGNL 60  
QY 61 ELTYLPTNASLFLDIOGVGVYVLAHNOVQVPLQRLRIYRGQLFEDNVALAVLNG 120  
Db 61 ELTYLPTNASLFLDIOGVGVYVLAHNOVQVPLQRLRIYRGQLFEDNVALAVLNG 120  
QY 121 DPLNNTPTVTGASPGGLRLQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNOLA 180  
Db 121 DPLNNTPTVTGASPGGLRLQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNOLA 180  
QY 181 LFLIDTNRSRACHPCSPMKGSRGWESSEDCQSLRTVACAGCARKGRLPTDCCHEQC 240  
Db 181 LFLIDTNRSRACHPCSPMKGSRGWESSEDCQSLRTVACAGCARKGRLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDFESHPNPEGRTYFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDFESHPNPEGRTYFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR 340  
Db 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR 340

## RESULT 4

US-09-854-356-6  
; Sequence 6, Application US/09854356  
; Patent No. US20020177567A1  
; GENERAL INFORMATION:  
; APPLICANT: Cheever, Martin A.  
; APPLICANT: Gheysen, Dirk  
; APPLICANT: Corixa Corporation  
; APPLICANT: SmithKline Beecham Biologicals S. A.  
; TITLE OF INVENTION: HER-2/neu Fusion Proteins  
; FILE REFERENCE: 014058-009810PC  
; CURRENT APPLICATION NUMBER: US/09/854,356  
; CURRENT FILING DATE: 2001-05-09  
; PRIOR APPLICATION NUMBER: US 09/493,480  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: US 60/117,976  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 6  
; LENGTH: 919  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein  
; OTHER INFORMATION: of ECD and PD of human HER-2/neu

US-09-854-356-6

Query Match 81.1%; Score 340; DB 9; Length 919;

Best Local Similarity 100.0%; Pred. No. 1.3e-302; Mismatches 0; Indels 0; Gaps 0;

Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 METALACRMGLLLALLPPGAASSTOYCTGTDMLRLPASPEHLMRLHLYOGCOVVOGNTL 60
DB 1 METALACRMGLLLALLPPGAASSTOYCTGTDMLRLPASPEHLMRLHLYOGCOVVOGNTL 60
OY 61 ELTYLPFNASLSFLQDIOEVQGYVLLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
DB 61 ELTYLPFNASLSFLQDIOEVQGYVLLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
OY 121 DPLNNTPTVYGASPGGLREQLRSLEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVYGASPGGLREQLRSLEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVACAGCARGKPLPTDCCHQC 240
DB 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVACAGCARGKPLPTDCCHQC 240
OY 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
DB 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
OY 301 YNLTSTDVGSCTLVCPHNOEYTAEDGTORCEKSKPCAR 340
DB 301 YNLTSTDVGSCTLVCPHNOEYTAEDGTORCEKSKPCAR 340
```

RESULT 5

US-09-769-508-2

Sequence 2, Application US/09769508

Patent No. US20020155527A1

GENERAL INFORMATION:

APPLICANT: STUART, SUSAN G.

APPLICANT: MONAHAN, JOHN J.

APPLICANT: LANGTON, BEATRICE CLAUDIA

APPLICANT: HANCOCK, MIRIAM E.C.

APPLICANT: CHAO, IORRINE A.

APPLICANT: BLUFORD, PETER

TITLE OF INVENTION: C-ERBB-2 EXTERNAL DOMAIN: GP75

FILE REFERENCE: BEBIO-111-C1

CURRENT APPLICATION NUMBER: US/09/769,508

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patentln Ver. 2.1

SEQ ID NO 2

LENGTH: 1255

TYPE: PRT

ORGANISM: Homo sapiens

US-09-769-508-2

Query Match 81.1%; Score 340; DB 9; Length 1255;

Best Local Similarity 100.0%; Pred. No. 1.7e-302;

Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 METALACRMGLLLALLPPGAASSTOYCTGTDMLRLPASPEHLMRLHLYOGCOVVOGNTL 60
DB 1 METALACRMGLLLALLPPGAASSTOYCTGTDMLRLPASPEHLMRLHLYOGCOVVOGNTL 60
OY 61 ELTYLPFNASLSFLQDIOEVQGYVLLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
DB 61 ELTYLPFNASLSFLQDIOEVQGYVLLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
OY 121 DPLNNTPTVYGASPGGLREQLRSLEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVYGASPGGLREQLRSLEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVACAGCARGKPLPTDCCHQC 240
DB 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVACAGCARGKPLPTDCCHQC 240
```

```
OY 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
DB 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
```

```
OY 301 YNLTSTDVGSCTLVCPHNOEYTAEDGTORCEKSKPCAR 340
DB 301 YNLTSTDVGSCTLVCPHNOEYTAEDGTORCEKSKPCAR 340
```

RESULT 6

US-09-854-356-1

Sequence 1, Application US/09854356

Patent No. US20020177567A1

GENERAL INFORMATION:

APPLICANT: Cheever, Martin A.

APPLICANT: Gheysen, Dirk

APPLICANT: Corixa Corporation

APPLICANT: SmithKline Beecham Biologicals S. A.

TITLE OF INVENTION: HER-2/neu Fusion Proteins

FILE REFERENCE: 014058-009810PC

CURRENT APPLICATION NUMBER: US/09/854,356

CURRENT FILING DATE: 2001-05-09

PRIOR APPLICATION NUMBER: US 09/493,480

PRIOR FILING DATE: 2000-01-28

PRIOR APPLICATION NUMBER: US 60/117,976

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 26

SOFTWARE: Patentln Ver. 2.1

SEQ ID NO 1

LENGTH: 1255

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: human HER-2/neu protein

NAME/KEY: DOMAIN

LOCATION: (1)..(653)

OTHER INFORMATION: extracellular domain (ECD)

NAME/KEY: DOMAIN

LOCATION: (676)..(1255)

OTHER INFORMATION: Intracellular domain (ICD)

NAME/KEY: DOMAIN

LOCATION: (990)..(1255)

OTHER INFORMATION: phosphorylation domain (PD)

NAME/KEY: DOMAIN

LOCATION: (990)..(1048)

OTHER INFORMATION: fragment of the phosphorylation domain, preferred

OTHER INFORMATION: portion (delta PD)

US-09-854-356-1

Query Match 81.1%; Score 340; DB 9; Length 1255;

Best Local Similarity 100.0%; Pred. No. 1.7e-302; Mismatches 0; Indels 0; Gaps 0;

Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 METALACRMGLLLALLPPGAASSTOYCTGTDMLRLPASPEHLMRLHLYOGCOVVOGNTL 60
DB 1 METALACRMGLLLALLPPGAASSTOYCTGTDMLRLPASPEHLMRLHLYOGCOVVOGNTL 60
OY 61 ELTYLPFNASLSFLQDIOEVQGYVLLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
DB 61 ELTYLPFNASLSFLQDIOEVQGYVLLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNG 120
OY 121 DPLNNTPTVYGASPGGLREQLRSLEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVYGASPGGLREQLRSLEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVACAGCARGKPLPTDCCHQC 240
DB 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVACAGCARGKPLPTDCCHQC 240
OY 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
DB 241 AACCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300
```

**Oy**      301 IYVLSITPVGSCILLCPPLHQETAEDGIQRCECKSKPPAR 340  
|||||  
|||||  
**Dd**    301 YNVLSITPDVGSCLLVCPLHNGETAADGTQRCCEKCSKPAPR 340

```

RESULT 7
US-09-930-125-2
: Sequence 2, Application US/09930125
: Publication No. US20020193329A1
: GENERAL INFORMATION:
: APPLICANT: Hand-zimmerman, Susan
: APPLICANT: Cheever, Martin A.
: APPLICANT: Foy, Teresa M.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Kalos, Michael D.
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Vedvick, Thomas S.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS
: OF HER-2/NEU-ASSOCIATED MALIGNANCIES
: FILE REFERENCE: 210121.544
: CURRENT APPLICATION NUMBER: US/09/930,125
: CURRENT FILING DATE: 2001-08-14
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 1255
: TYPE: PRT
: ORGANISM: Homo sapien
: US-09-930-125-2

```

RESULT 8  
US-09-441-411-6  
: Sequence 6, Application US/09441411  
: Publication No. US2003008342A1  
: GENERAL INFORMATION:  
: APPLICANT: Scholler, Nathalie B.  
: APPLICANT: Disis, Mary L.  
: APPLICANT: Hellstrom, Ingegerd  
: APPLICANT: Hellstrom, Karl Erik  
: TITLE OF INVENTION: SURFACE RECEPTOR ANTIGEN VACCINES  
: FILE REFERENCE: 730033.409  
: CURRENT APPLICATION NUMBER: US/09/441.411  
: CURRENT FILING DATE: 1999-11-16

```

: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 1255
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-441-411-6

```

```

RESULT 9
US-09-811-123-9
: Sequence 9, Application US/09811123
: Patent No. US20020001587A1
: GENERAL INFORMATION:
: APPLICANT: Sharon Erickson
: APPLICANT: Ralph Schwall
: APPLICANT: Mark Sliwowski
: TITLE OF INVENTION: METHODS OF TREATMENT USING ANTI-EPDB
: TITLE OF INVENTION: ANTIBODY-MAYTANSINOID CONJUGATES
: FILE REFERENCE: GENEENT.073A2
: CURRENT APPLICATION NUMBER: US/09/811.123
: CURRENT FILING DATE: 2001-03-16
: PRIOR APPLICATION NUMBER: 60/238,327
: PRIOR FILING DATE: 2000-10-05
: PRIOR APPLICATION NUMBER: 09/602,530
: PRIOR FILING DATE: 2000-06-23
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9
: LENGTH: 1255
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-811-123-9

      81.1%; Score 340; DB 10; Length 1255;
      Best Local Similarity 100.0%; Pred. No. 1.7e-302;
      Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY      1 MELAALCRWGLLALLPPGCASTGYCTGDMKRLRPASPFTHLDMRLHYGCGVQVGNL 60
DB      1 MELAALCRWGLLALLPPGCASTGYCTGDMKRLRPASPFTHLDMRLHYGCGVQVGNL 60

QY      61 ELLTYPTNASTLSFDODIDEGVGYVLIANHQVPLQRLRIYRGTOLEFDNATALAVDNG 120
DB      61 ELLTYPTNASTLSFDODIDEGVGYVLIANHQVPLQRLRIYRGTOLEFDNATALAVDNG 120

```







US-09-821-883-4

Query Match 69.7%; Score 292; DB 10; Length 697;

Best Local Similarity 100.0%; Pred. No. 7.6e-259; Mismatches 0; Indels 0; Gaps 0;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 19 GAASIVCTGTDMKRLPASPETHLMDLRHLYOGCVOVGNLELYLPTNASLSFLQDIO 78
    |||||||
Db 38 GAASIVCTGTDMKRLPASPETHLMDLRHLYOGCVOVGNLELYLPTNASLSFLQDIO 97
    |||||||
QY 79 EVQGVYLLAHNGVROVPLQRLRIVRGTQLEEDNYALAVLDNGDPLNNTPTVYGASPGGLR 138
    |||||||
Db 98 EVQGVYLLAHNGVROVPLQRLRIVRGTQLEEDNYALAVLDNGDPLNNTPTVYGASPGGLR 157
    |||||||
QY 139 ELQLRSLEILKGVLIQRNPOLCYODTILMKDIFHKNNQALATLIDTNRSRACHPCSPM 198
    |||||||
Db 158 ELQLRSLEILKGVLIQRNPOLCYODTILMKDIFHKNNQALATLIDTNRSRACHPCSPM 217
    |||||||
QY 199 CKGSRGWSESSDDCSLRTVVCAGGCARCKGPLPTDCHHEQCAAGCTGPKHSDCLACLF 258
    |||||||
Db 218 CKGSRGWSESSDDCSLRTVVCAGGCARCKGPLPTDCHHEQCAAGCTGPKHSDCLACLF 277
    |||||||
QY 259 NMSGICELHCPALVYNTDTFESMPNPEGRTYTFGASCVTACPYNYLSTDVGS 310
    |||||||
Db 278 NMSGICELHCPALVYNTDTFESMPNPEGRTYTFGASCVTACPYNYLSTDVGS 329
    |||||||
```

Search completed: March 4, 2003, 12:54:28  
Job time : 20.6687 secs

